

MILIK FAKULTAS HUKUM
UNIVERSITAS AIRLANGGA

PENGETAHUAN DASAR

PENELITIAN DAN STATISTIK

K. PENELITIAN

KKS

KK

001.4

Jaw

P



PENATARAN UNTUK PERGURUAN TINGGI
DI JAWA TIMUR
OLEH
PIPR JAWA TIMUR
1976

KATA PENGANTAR

Buku PENGETAHUAN DASAR PENELITIAN DAN STATISTIK ini merupakan suntingan naskah-naskah kertas kerja Penataran Staf Pengajar Perguruan Tinggi di Jawa Timur yang diselenggarakan oleh Pusat Ilmiah dan Pengembangan Regional Jawa Timur tahun 1976.

Dengan jumlah terbatas suntingan pertama naskah-naskah tersebut telah diterbitkan pada tahun 1976. Cara penyajian naskah-naskah tersebut pada suntingan pertama begitu lugas, sehingga terdapat hal-hal yang dirasa kurang efektif sebagai suatu pustaka. Oleh karena itu dengan tidak mengurangi kandungan maknanya suntingan naskah-naskah tersebut perlu diperbaiki.

Tim editor telah mencoba memperbaiki suntingan naskah-naskah tersebut dengan mengintegrasikan rangkaian naskah dan lampiran lembaran tugas, sehingga pada setiap bab tersaji suatu uraian yang terintegrasi. Dengan cara ini maka suntingan naskahnya dapat digunakan oleh staf pengajar lain yang pada kesempatan penataran tersebut tidak sempat mengambil bagian aktif sebagai peserta. Urutan bab-babnya telah dirakit kembali sehingga rangkaian informasinya serasi untuk dipahami.

Penerbitan buku ini dan penyebarannya secara luas bertujuan untuk memenuhi harapan tumbuhnya kegairahan kegiatan riset sebagai perwujudannya ta Dharma Penelitian sebagai salah satu misi Perguruan Tinggi. Harapan tersebut sesuai dengan maksud yang tersurat pada Kebijakan Dasar Pengembangan Pendidikan Tinggi Menteri Pendidikan dan Kebudayaan tentang Dasar dan Arah Pembinaan dan Pengembangan Perguruan Tinggi, yang antara lain mengungkapkan:

- (a) bahwa kegairahan riset perlu ditumbuhkan, sehingga Pendidikan Tinggi dapat berpartisipasi dalam perbaikan serta pengembangan ilmu pengetahuan dan penerapannya;
- (b) bahwa Pendidikan Tinggi harus merupakan bagian integral dari usaha-usaha pembangunan baik nasional maupun regional, serta merupakan penghubung antara dunia ilmu pengetahuan, teknologi dan kebutuhan masyarakat.

Permintaan bantuan konsultasi oleh masyarakat, pemerintah, dan swasta kepada Perguruan Tinggi yang makin meningkat perlu dipenuhi melalui peningkatan kemampuan Perguruan Tinggi di bidang penelitian. Mengingat heterogenitas kemampuan Perguruan Tinggi untuk memenuhi permintaan tersebut maka perlu dipupuk kerjasama ilmiah antar Perguruan Tinggi. Kerjasama ilmiah yang menyangkut kegiatan penelitian memerlukan bahasa berfikir yang sama melalui keseragaman metode.

Menyadari bahwa betapa pun bervariasi, penelitian pada dasarnya mempunyai dasar-dasar kesamaan metode. Oleh karena itu dengan penyebaran buku ini sebagai langkah pertama menuju ke bentuk kerjasama yang lebih maju merupakan usaha yang patut ditempuh.

Naskah aseli yang tersunting dalam buku ini merupakan sumbangan berharga staf pengajar dari Institut Teknologi Sepuluh Nopember Surabaya; Institut Keguruan dan Ilmu Pendidikan Surabaya; dan Universitas Airlangga Surabaya. Dengan tidak mengurangi potensi yang tersedia di Perguruan Tinggi lain di Jawa Timur sumbangan naskah dari ketiga perguruan tinggi tersebut yang nama penulisnya tersurat pada buku ini merupakan batu pertama dari usaha kerjasama ilmiah perguruan tinggi yang patut dihargai.

Semoga buku ini mencapai harapan kita bersama dalam memenuhi tuntutan seperti terkandung pada Tri Dharma Perguruan Tinggi, khususnya Dharma Penelitian.

Surabaya, 31 Oktober 1978.

Ketua Badan Pekerja PIPR/RSDC
Jawa Timur

Prof. Dr. A.A. Loedin
NIP.130162036.

DAFTAR ISI

	Halaman
Kata Pengantar	ii
1. Pengertian Dasar Tentang Metodik Penelitian	1
2. Statistik Deskriptip	24
3. Statistik Analitis	42
4. Pengumpulan Data	59
5. Teknik Sampling	66
6. Pengolahan Data	79
7. Dokumentasi dan Publikasi Penelitian	92
8. Penyusunan Usul Proyek	101

1. PENGERTIAN DASAR TENTANG METODIK PENELITIAN

Scientific research is a systematic, controlled, empirical and critical investigation of hypothetical propositions about the presumed relations among natural phenomena.

Dengan Tridharma tugas-tugas utama Perguruan Tinggi telah jelas digaris-kan. Tiga Dharma ini biasanya disebut dalam urutan Pendidikan, Peneliti-an, dan Pengabdian pada Masyarakat.

Science is the making of knowledge and not knowledge as such, so that it has become more nearly equated with research.

Kalau pilihan deretan ini ditinjau lebih mendalam, maka tampak bahwa se-benarnya susunan deretan ini tak benar.

Perguruan Tinggi ialah tempat di mana ilmu dikembangkan, pengembangan ilmu ini dilakukan dengan penelitian. Hasil penelitian yang berupa tam-bahan pada ilmu pengetahuan dapat disumbangkan kembali pada ilmu penge-tahuan sendiri atau diamalkan pada umat manusia. Pengamalan ilmu ini di-sebut Pengabdian pada Masyarakat.

Untuk dapat menyampaikan ilmu pengetahuan secara baik kepada masyarakat diperlukan orang-orang yang mampu meneruskannya. Untuk ini perlu diada-kan pendidikan. Kalau kita mengikuti jalan pikiran tersebut di atas, ma-ka susunan deretan Dharma yang sewajarnya ialah Penelitian, Pendidikan dan Pengabdian pada Masyarakat.

Jalan pikiran ini jelas berlaku di negeri-negeri yang maju di mana Per-guruan-perguruan Tinggi sudah mendapat tempat dan fungsi yang sebenarnya. Di Indonesia, yang masih tergolong dalam kelompok negara-negara yang se-dang berkembang, hal ini agak berlainan.

Salah satu definisi negara yang berkembang, yang sekarang dianggap pa-ling tepat, ialah : Negara yang masih berkembang ialah negara yang masih harus mengimport "science and technology".

Jika definisi ini benar maka perkembangan ilmu pengetahuan menjadi tugas utama Perguruan Tinggi dan dengan demikian deretan yang disebut dalam kalimat di atas dapat dibenarkan.

Di Perguruan-perguruan Tinggi di Indonesia sudah terdapat taraf perkembangan di mana Penelitian mengambil peranan yang makin lama makin penting. Sebagian ini ialah akibat dari kematangan Perguruan Tinggi, sehingga dengan sendirinya unsur-unsur Tridharma tersusun menurut deretan yang sebenarnya.

Selain dari ini juga tampak bahwa masyarakat, pemerintah pusat, pemerintah daerah, dan instansi-instansi partikelir, makin lama makin banyak datang ke Perguruan Tinggi sebagai tempat bertanya dan berkonsultasi. Sikap pemerintah pusat dan daerah tersebut di atas sebenarnya sudah lama tampak, tetapi kegiatan nyata bertambah sesudah anjuran Bapak Presiden di Bandung pada tahun 1970 kepada instansi-instansi pemerintah untuk lebih mengeratkan hubungan dengan Perguruan Tinggi di daerah masing-masing. Hal ini dapat ditemukan kembali dalam keputusan Menteri Dalam Negeri No. 2/1971.

Telah disebut bahwa Perguruan Tinggi ialah tempat di mana Ilmu Pengetahuan (Science) dikembangkan, dipelihara, dan diamalkan.

Setiap buku yang membicarakan ilmu pengetahuan, mulai dengan menjelaskan dan merumuskan apakah yang dimaksud dengan Ilmu Pengetahuan ini.

Ternyata setiap penulis dan setiap buku menyebut perumusan lain. Tetapi kalau diteliti, maka pada semua perumusan yang sepintas lalu tampak berlainan itu ada persamaan-persamaan.

Yang dianggap sebagai ilmu pengetahuan terdiri dari dua unsur:

1. sejumlah pengetahuan (Body of knowledge).
2. metodik untuk mengembangkan ilmu pengetahuan ini.

Ilmu Pengetahuan ialah suatu pengertian dinamis dan tak dapat dilihat lepas dari lingkungan sekitarnya dan dari sejarahnya yang lampau. Inilah yang menyebabkan kesukaran dalam merumuskan definisi yang memuaskan.

Kesukaran perumusan yang sama juga terdapat untuk hal-hal yang semacam, umpamanya kesenian (art). Yang disebut di atas terutama berlaku "the body of knowledge" tetapi tak begitu berpengaruh pada metodik.

Dalam abad ke-20 ini ilmu pengetahuan cepat sekali berkembang, dengan suatu puncak sewaktu dan sesudah Perang Dunia ke-2.

Dalam perkembangan yang makin lama makin bertambah cepat ini, tampak bahwa ilmu pengetahuan dipecah menjadi bagian-bagian yang terpisah (spesialisasi). Bagian-bagian ini kemudian dipecah-pecah lagi dan seterusnya. Keadaan sekarang ini ialah bahwa ilmu pengetahuan telah terbagi dalam bidang-bidang yang begitu kecil sehingga sarjana-sarjana dari Fakultas, bagian, atau sub bagian yang berlainan tak dapat berbicara lagi satu sama lain. Akibat pemisahan ini sangat terasa. Perkembangan-perkembangan bagian yang terpecah-pecah ini terus berjalan, tetapi ilmu pengetahuan sebagai suatu kegiatan utama manusia sukar berkembang. Hal ini sudah dirasakan dan menjadi alasan terbentuknya "Advance Colleges" dan sebagainya, di mana para sarjana yang terkemuka dari berbagai bidang dapat bertemu lagi. Untunglah bahwa dalam suasana perpecahan ini tampak satu titik terang.

Metodik ilmu pengetahuan tak banyak berubah dan dalam garis-garis besarnya sama untuk semua bidang ilmu pengetahuan. Oleh karena itu di sini terutama dibicarakan ialah metodiknya dan cara penggunaannya.

Selama manusia hidup di dunia ini pengetahuan dan pengalamannya selalu bertambah. Ini disebabkan oleh dua dorongan yang kuat. Pertama ialah usaha manusia untuk memperbaiki hidupnya. Dengan cara ini ditemukan api, roda, dan layar. Untuk pedoman kapal-kapal yang berlayar jauh dari pantai perlu ditemukan pedoman-pedoman yang dapat digunakan untuk menentukan di mana kapal berada dan di mana pelabuhan sasarannya. Dengan demikian bintang-bintang diperhatikan dan dipelajari sehingga 6 abad sebelum Masehi orang Babylonia sudah dapat meramalkan gerhana matahari dan bulan. Untuk dapat memperhitungkan kapan dan di mana sungai Nil akan meluap oleh orang Mesir kuno dilahirkan dasar-dasar kalender dan geometri.

Dorongan kedua ialah sifat manusia yang ingin mengetahui segala sesuatu. Pada permulaan manusia hanya menambah pengetahuan dan pengalamannya dengan mencoba-coba saja (Trial and Error method). Cara ini sering mengecewakan dan manusia mencari pegangan-pegangan lain untuk menentukan sesuatu. Pegangan-pegangan yang dipakai dan cara-cara yang dikembangkan ialah berdasarkan:

1. Kewibawaan.

Untuk persingkatan waktu dan untuk mempermudah diri sendiri dalam mengambil keputusan, manusia meminta pendapat orang-orang di kalangan-nya yang dianggap mempunyai pengetahuan dan pengalaman banyak.

2. Pengalaman pribadi.

Cara ini dapat dilihat sebagai suatu kemajuan dalam cara pemikiran. Di sini manusia tak tunduk lagi pada kewibawaan tetapi ingin mengetahui sendiri. Meskipun cara ini sudah berarti langkah maju, masih tetap terdapat kekurangan-kekurangan yang perlu diperhatikan supaya kita tidak tersesat.

Kekurangan-kekurangan ini ialah:

- pengalaman manusia terbatas
- menghilangkan hal-hal yang tak sesuai dengan pendapatnya
- terlalu cepat percaya
- mengambil keputusan yang sesuai dengan pemikiran sendiri (pre-judice).

3. Deduksi.

Perbaikan berikut ialah dari Aristoteles yang mengembangkan syllogisme, suatu pemikiran dengan menggunakan premises yang memungkinkan pengujian kebenaran dari kesimpulan yang ditarik.

4. Induksi.

Dapat dianggap sebagai penyempurnaan dari pemikiran deduksi dengan menambah usaha mencari bukti-bukti bahwa premises yang dipakai benar dan dapat dipercaya.

Perlu diketahui bahwa cara pemikiran dibagi dalam tiga macam:

A. *Induksi sempurna.*

Di sini setiap unsur (data) harus diperiksa satu persatu, sehingga terjadi yang disebut Complete Enumerration. Hal ini berupa sesuatu yang jarang sekali dapat dilaksanakan.

B. *Induksi tak sempurna.*

Di sini hanya sebagian bahan diperiksa untuk dapat mengambil suatu kesimpulan yang menyeluruh.

Persoalan yang timbul nanti masih akan dibahas pada pembicaraan

Sampling.

C. Induksi dari Bacon.

Francis Bacon (1561 - 1626) mempunyai keberatan terhadap pendekatan deduksi yang bertitik tolak dari premises yang tampaknya jelas atau berdasarkan pada kewibawaan. Dia ingin bahwa peneliti memperhatikan sendiri alam sekitarnya untuk mengambil kesimpulan.

Pemikiran Bacon inilah yang menjadi salah satu dasar perkembangan pemikiran yang disebut scientific method.

Berpikir, sebagai kemampuan manusia yang tertinggi, tak boleh dilihat lepas dari perkembangan manusia sebagai makhluk di dunia ini. Dari jaman Aristoteles sampai sekarang, oleh para sarjana telah diterima bahwa tak ada garis pemisah yang jelas antara atom sederhana dan manusia sebagai makhluk yang tertinggi di dunia ini atau antara benda mati dan makhluk hidup. Kalau hidup dirumuskan dalam istilah-istilah biokimia, maka hidup ialah suatu rentetan proses-proses biokimia yang berlangsung pada suhu yang rendah dengan katalysasi dibantu oleh ensim-ensim.

Pada permulaan di dunia terdapat lautan dengan terlarut di dalamnya gabungan-gabungan kimia anorganik. Pada waktu ini kegiatan yang terjadi dapat dilagangkan sebagai berikut:

AKSI → REAKSI

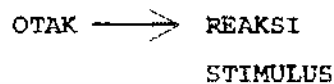
Menurut teori Evolusi Biokimia dari Oparin dan Haldane (1920) dari bahan-bahan anorganik terbentuk bahan-bahan organik. Dengan terdapatnya bahan organik dapatlah terbentuk zat putih telur dan dengan demikian unsur-unsur untuk membangun sel sudah tersedia. Dengan terbentuknya DNA enersi dapat dibangkitkan dan hidup dalam bentuk yang paling primitif dapat dimulai.

Bagan sederhana AKSI → REAKSI, untuk sel dan binatang yang terdiri dari beberapa sel menjadi STIMULUS → REAKSI.

Pada binatang yang lebih tinggi perkembangannya dan pada manusia tersusun sel-sel dengan fungsi-fungsi khusus antara lain urat-urat syaraf yang kemudian menjadi otak.

Seluruh kegiatan manusia disalurkan melewati otak ini. Namun demikian pada manusia masih dapat dilihat bahwa untuk proses penyelaan hidupnya ia masih jatuh kembali pada bagan primitif itu.

Seorang yang menyentuh api akan menarik kembali tangannya sebelum ia dapat menyadari dengan otaknya, bahwa ada rasa nyeri dan suatu keadaan yang membahayakan. Tindakan semacam ini disebut *Reflex*. Kalau pada manusia di dalam bagan tersebut di atas disisipkan otak, maka terdapat :



Bagan yang sangat sederhana ini dapat diperbaiki dan disempurnakan. Stimulus sebenarnya ialah suatu pengaruh dari luar yang dialami dan dirasakan oleh organisme tersebut.

Manusia dalam lingkungan hidupnya menggunakan pancainderanya untuk menerima stimulus. Ia akan melihat, meraba dan sebagainya sehingga dia dapat mengenal dan memeriksa sebagian dari lingkungan sekitarnya. Dengan demikian perkataan stimulus di sini berarti

Observasi (I)

Hal-hal yang diperoleh pada observasi disalurkan liwat otaknya dan dengan kemampuannya diadakan registrasi. Pada suatu waktu di alam yang diobserveer ini akan tampak sesuatu yang ganjil atau menghe-rankan dan pada manusia tersebut timbulah pertanyaan. Dia akan merumuskan pertanyaan yang timbul. Dengan lain perkataan dia meng-adakan *Problem Statement* (II)

Menggunakan pengalaman-pengalaman dan dengan berpikir (analogy and imagination), dia akan mencoba menjawab pertanyaan ini, dengan la-in perkataan disusun suatu *Hypothesis* (III)

Tentu saja karena pemikirannya yang kritis dia ingin membuktikan bahwa jawabannya (hipotesanya) ialah benar dan untuk pembuktian ini dia akan menyusun suatu rencana kerja, inilah yang disebut

Design (IV)

Dalam bagan sederhana di atas sekarang kita sampai pada REAKSI atau tindakan-tindakan rencana kerja yang telah disusun dijalankan, *Execution* (V)

Hasil pekerjaan ini (*Result VI*) perlu disebar luaskan ke dunia il-mu pengetahuan dapat berkembang *Documentation* (VII)

Yang disebut dari angka I sampai VII ialah tidak lain dari tindakan-tindakan/langkah-langkah yang digunakan dalam Metodik Ilmu Pengetahuan:

- I. Observation (Pengamatan)
- II. Problem statement (Perumusan Persoalan)
- III. Hypothesis (Hipotesa)
- IV. Design (rencana kerja)
- V. Excecution (pelaksanaan)
- VI. Result (hasil)
- VII. Documentation (Dokumentasi)

Jika jalan ini sudah ditempuh maka ceritera akan berulang. Hasil penelitian ini akan ditinjau kembali dan akan direcycle lagi. Tak ada suatu perkembangan ilmiah yang tak berdasarkan perkembangan yang lebih dulu. Karena kita tiap kembali ke asal untuk kemudian maju lagi maka kegiatan ini mendapat nama research (Penelitian). Kegiatan ini tak berjalan dalam suatu lingkaran tetapi berupa suatu spiral. Setiap putaran kita berada di tingkat yang lebih tinggi. Kenaikan-kenaikan ini berarti perkembangan manusia dan ilmu pengetahuannya.

*Science begins with the observation
of selected parts of nature.*

I. Pengamatan.

Setiap kegiatan ilmiah dimulai dengan pengamatan lingkungan sekitarnya. Lingkungan ini dapat luas atau terbatas. Pengamatan dapat di lingkungan luas, seperti kadang-kadang dilakukan pada penelitian soaial dengan turun ke lapangan, tetapi dapat juga dilakukan dalam lingkungan sangat terbatas; umpamanya pengamatan nucleus sel dengan mikroskop elektron atau dengan cara biokimia mikro. Pengamatan-pengamatan ini ialah pengamatan langsung oleh si peneliti sendiri, tetapi cara yang lain ialah membaca apa yang telah dilakukan oleh sarjana-sarjana lain dengan menggunakan perpustakaan.

Pengamatan Perpustakaan juga dapat menghindari suatu duplikasi, yaitu usaha menjawab pertanyaan yang sebenarnya sudah terjawab. Hal ini akan memberi kekecewaan dan penghamburan uang, waktu dan tenaga manusia.

Six hours in the library may save six months in the laboratory.

Dari pengalaman sehari-hari diketahui bahwa seorang hanya mencari sesuatu dengan baik, kalau ia tahu benar-benar apa yang dicari dan ia juga kenal tempat di mana benda yang dicari berada. Meskipun seorang melihat lingkungan sekitarnya dengan saksama untuk jangka waktu yang cukup lama, kalau ia tidak siap sebelumnya, maka ia tak akan menemui sesuatu.

Untuk mengadakan pengamatan yang baik diperlukan dua hal, yaitu pikiran yang siap-siaga (PREPARED MIND) dan cara berpikir yang teratur dan terlatih (SCIENTIFIC MIND).

Pikiran seorang harus disiapkan untuk dapat menampung hal/keadaan yang menyolok di lingkungan yang diamati. Untuk ini diperlukan pengetahuan umum yang luas; yang dimaksud ialah bahwa ia juga mempunyai pengetahuan dalam bidang-bidang ilmu pengetahuan lain kecuali keahliannya sendiri.

Tentu saja pengetahuan ini tak akan mendalam, tetapi dengan mengetahui sedikit banyak dari bidang-bidang lain maka penglihatannya akan lebih lebar dan daya tangkapnya juga akan bertambah.

Kecuali ini juga diperlukan pengetahuan yang mendalam di satu atau beberapa bidang ilmu pengetahuan.

Chance favoured a prepared mind.

Contoh-contoh dalam sejarah ilmu pengetahuan cukup banyak. Penemuan perhitungan gaya berat oleh Sir Isaac Newton (1642-1727), perumusan dalil Archimedes (287-212 B.C.), penemuan obat penicillin oleh Sir Flemming; adalah contoh-contoh bahwa dalam hal-hal yang sudah sering dilihat oleh banyak orang dapat ditemukan sesuatu yang baru oleh seorang yang memiliki

Yang juga dibutuhkan oleh setiap orang yang ingin melakukan Penelitian ialah suatu "Scientific Mind." Seorang Ilmuwan harus dapat berpikir:

- a. Obyektif : artinya dapat berpikir tanpa dipengaruhi perasaan-perasaan atau kemauan-kemauan diri sendiri
- b. Independent : tak mau dipengaruhi oleh orang-orang lain, keadaan politik atau apapun yang ada di sekitarnya.
- c. Menolak suatu kewibawaan dalam ilmu pengetahuan.

There is no authority in science

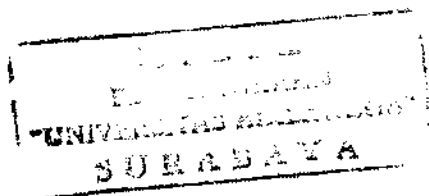
Penerimaan kewibawaan dalam ilmu pengetahuan dapat menyesatkan dan menghentikan perkembangannya. Pengaruh gereja Katolik terhadap Galileo Galilei (1564 - 1642) di bidang Astronomi dan Cosmografi dan terhadap Petucci (1961) di bidang Biologi. Pengaruh perorangan banyak terdapat dalam sejarah, begitu juga pengaruh politik.

Dunia ilmu pengetahuan kadang-kadang dapat begitu tersesat, sehingga terdapat pengakuan yang menyeluruh tentang sesuatu yang tak benar.

*Universal agreement can be obtained
about the untruth.*

Seorang ilmuwan menerima hal-hal yang dia amati sendiri secara langsung atau tidak langsung, atau sesuatu yang dilihat oleh sarjana lain, asal cara melihat dan menilai sarjana tersebut memenuhi syarat dan dapat dipercaya.

Ternyata bahwa terdapat suatu bidang ilmu pengetahuan yaitu sexuologi, yang hanya didasarkan atas informasi yang diperoleh dari orang-orang yang bersangkutan. Harus diketahui bahwa orang yang bersangkutan ialah seorang awam dalam bidang sexuologi dan mungkin tak dapat berpikir sendiri tanpa di-



II. Perumusan Persoalan.

Pertama-tama yang perlu dijelaskan ialah bahwa yang dimaksud dalam ilmu pengetahuan dengan suatu persoalan, berlainan dengan yang di dalam percakapan sehari-hari disebut suatu persoalan. Di dalam hidup sehari-hari yang biasanya dimaksud dengan persoalan ialah suatu kesukaran, keganjilan, atau sesuatu yang menimbulkan "pusing kepala".

Untuk para ilmuwan, persoalan mempunyai arti yang khas dan harus memenuhi beberapa persyaratan, yaitu:

1. menunjukkan hubungan antara dua variabel
2. persoalan harus ditegaskan dalam bahasa yang jelas dan untuk mudahnya dituangkan dalam kalimat pertanyaan.
3. persoalan harus memungkinkan pengukuran empiris.

Tiga point ini perlu dijelaskan.

ad. 1. Pengertian variabel.

Ilmu pengetahuan memakai dasar bahwa dalam alam semesta di mana kita berada terdapat suatu ketertiban.

There is some order in nature.

Artinya, apa yang kemarin terjadi, hari ini juga akan terjadi dalam keadaan sama. Tetapi untuk dapat menjamin bahwa yang timbul akan sama seperti dulu, semua faktor harus sama. Umpamanya, supaya air mendidih dalam waktu yang sama seperti kemarin, harus diatur dan dipastikan bahwa semua faktor sama seperti kemarin. Umpamanya jumlah air, suhu air awal, kejernihan air, bentuk dan bahan panci, macam api atau alat pemanas, tekanan udara, suhu sekitarnya, dan sebagainya. Faktor-faktor ini tersebut disebut variabel.

Dalam setiap penelitian harus ditentukan variabel yang diketahui dan apakah ada kemungkinan terdapat variabel yang tak diketahui. Lagi pula apakah variabel-variabel itu dapat diatur dan dikontrol, seperti sering terjadi dalam laboratoria, atau tak terkontrol.

Yang terakhir ini hampir selalu terdapat dalam penelitian yang meliputi makhluk hidup, umpamanya manusia. Persoalan ini tentu saja lebih jelas lagi dalam bidang-bidang ilmu sosial-budaya.

ad.3. Hasil akhir suatu penelitian selalu berarti suatu pengukuran.

Science is enumeration.

Sebelum dapat dibicarakan pengertian pengukuran, perlu dulu dimiliki pengertian tentang penggunaan dan arti angka-angka.

Angka tak lain dari suatu simbol yang dapat dipakai dengan empat cara dan artinya tergantung pada cara penggunaan.

Empat cara itu ialah dalam:

1. Skala Nominal
2. Skala Ordinal
3. Skala Interval
4. Skala Ratio.

Ada dua istilah yang perlu diketahui dan dibedakan, yaitu arithmetization dan measurement.

Arithmetization ialah pemberian nomor pada suatu obyek, kejadian atau sifat (1, 2, 3, dan 4,).

Measurement hanya pemberian angka dalam penggunaan ukuran yang tetap (3 dan 4).

Dalam ilmu pengetahuan dua hal ini tak dibatasi begitu jelas, sehingga measurement dipakai untuk yang sebenarnya arithmetization.

Pembicaraan ini akan menggunakan bagan di halaman 14.

TABLE A CLASSIFICATION OF SCALES OF MEASUREMENT

SCALE	BASIC EMPIRICAL OPERATIONS	MATHEMATICAL GROUP STRUCTURE	TYPICAL EXAMPLES
Nominal	Determination of equality	Permutation group $x^i = f(x)$ Where $f(x)$ means any one-to-one substitution	"Numbering" of football players. Assignment of type or model numbers to classes.
Ordinal	Determination of greater or less	Isotonic group $x^i = f(x)$, Where $f(x)$ means any increasing monotonic function.	Hardness of minerals, street-numbers, grades of leather, lumber, wool, etc. Intelligence test raw scores.
Interval	Determination of the equality of intervals or of differences	Linear or affine group $x^i = ax + b$, $a > 0$	Temperature (Fahrenheit or Celsius) Position Time (calendar) Energy (potential) Intelligence test "standard scores" (?).
Ratio	Determination of the equality of ratios	Similarity group $x^i = cx$, $c > 0$	Numerosity, length, density, work, time intervals, etc. Temperature (Rankine or Kelvin) Loudness (sound) Brightness (light)

ad. 1. Skala Nominal.

Dalam kata nominal tersimpan kata asal dari perkataan nama. Memang pada skala nominal ini, angka hanya berarti pemberian nama atau tanda sehingga benda tersebut dengan mudah dapat di kenal dan ditemukan kembali. Tak akan berubah sesuatu, kalau umpamanya dipakai huruf sebagai pengganti angka-angka tersebut.

Skala nominal bertitik tolak dari pengertian bahwa antara unsur-unsur yang diberi nomor tak ada perbedaan.

*Contoh: A. Pemain-pemain regu sepak bola di-beri nomor*di belakang bajunya.*

Nomor ini tak berarti bahwa yang memakai nomor satu ialah yang paling dahulu. Nomor ini hanya dipakai sebagai pengenalan saja.

B. Penggunaan angka pada Dewey Decimal System, sistem yang digunakan di-perpustakaan-perpustakaan.

C. Nomor mobil

D. Nomor induk pegawai

ad. 2. Skala Ordinal.

Dalam perkataan ordinal tersimpan pengertian orde, suatu aturan atau ketertiban. Kalau skala nominal menganggap semua unsur sama, pada skala ordinal diakui perbedaan dengan arti bahwa satu unsur lebih besar atau kecil dari yang lain.

Atas dasar kriteria penilaian tertentu, unsur disusun dan diberi nomor.

Contoh: A. Nomor-nomor rumah. Kalau di jalan terdapat rumah nomor 1 dan no.7

maka jelaslah bahwa di antaranya akan terdapat rumah-rumah no.3 dan 5. Tak berarti bahwa rumah - rumah no. 1, 2, 3, 5, dan 7 itu sama bentuk, sama besar, atau jarak antara rumah-rumah sama.

B. Skala dari Mohs. Dalam menentukan kekerasan batu dibuat suatu skala dari 1 sampai 10.

10 yang paling keras, yaitu intan, 1 yang paling lunak ialah talk.

Penggunaan skala ini ialah bahwa yang lebih keras dapat menggores yang lebih lunak, dan tak sebaliknya. Dengan mencoba-coba dapat ditentukan kekerasannya. Jelas bahwa yang berangka 10 lebih keras dari angka 9, tetapi tak berarti bahwa 10 dan 9 sama dengan 9 dan 8.

ad. 3. Skala Interval.

Di sini telah dicapai persamaan perbedaan.

Contoh: Pengukuran suhu dengan cara Celsius dan Fahrenheit. Jarak antara derajat satu dengan yang lain sama pada pengukuran Celsius maupun Fahrenheit. Tetapi karena angka 0 bukan 0 murni dan perbedaan dibuat atas perjanjian, maka Celsius tak dapat disamakan dengan Fahrenheit dan tak dapat dikatakan bahwa 30°C . itu dua kali lebih panas dari 15°C .

ad. 4. Skala Ratio.

Pengukuran dengan skala ratio menggunakan suatu ukuran tetap yang diketahui dan diakui. Pemindahan dari satu skala ratio ke yang lain dapat dicapai dengan mengalikan dengan angka tertentu.

Contoh: meter, gram, inch, foot, dan yang lain. Kalau suatu meja diukur dan terdapat panjang 3 meter, maka ini tak lain artinya bahwa meja tersebut 3 kali lebih panjang dari standar meter yang disimpan di Paris. Dengan lain perkataan yang dilakukan tak lain dari membandingkan.

Sering kali pada penelitian diadakan perbandingan antara bahan yang diperiksa dengan bahan yang telah dikenal dan tak akan berubah, bahan kedua ini disebut kontrol. Sebaliknya kontrol ini diketahui dan dikenal oleh ilmuwan lain, sehingga kalau dianggap perlu oleh mereka dapat digunakan kontrol yang sama.

Kalau suatu kontrol sudah diketahui dan mendapat pengakuan maka ini disebut suatu standar.

Di atas sudah dijelaskan bahwa pengertian "persoalan" dalam ilmu pengetahuan berlainan dengan yang disebut "persoalan" dalam pembicaraan sehari-hari. Persoalan ialah unsur yang terpenting dalam penelitian, bersama dengan hipotesa dan percobaan (experiment). Selama tiga unsur ini tak terdapat maka kegiatan tersebut belum dapat disebut penelitian.

Sebelum tahun 1966 kegiatan penelitian di Indonesia dianggap sebagai barang mewah yang tak begitu perlu dan hanya dilakukan oleh beberapa orang dengan hampir tak tersedia dana. Sekarang dengan perubahan iklim di Negara kita, tampak bahwa penelitian menjadi suatu "mode". Setiap instansi melakukan penelitian dan segala sesuatu perlu atau tak perlu diteliti. Dengan perkembangan ini maka sering kegiatan yang tak memenuhi syarat juga disebut penelitian.

Pada taraf pengamatan atau pada taraf perumusan persoalan, kadang-kadang ingin diketahui keadaan sebenarnya dari sesuatu yang berhubungan dengan penelitian itu. Akan dikumpulkan data yang diperlukan untuk dapat memberi gambaran yang dibutuhkan. Kegiatan ini, di mana data dikumpulkan dan kemudian disusun secara teratur untuk mendapatkan gambaran sesuatu keadaan, disebut survey. Survey sendiri tak merupakan sumbangan pada ilmu pengetahuan dan tak akan dapat memecahkan suatu persoalan. Survey hanya suatu taraf dalam kegiatan penelitian yang kompleks itu.

"If Research were limited simply to gathering so-called facts, scientific knowledge could not advance. Many people think that science is basically a fact-gathering activity".

Kalau persoalan sudah dirumuskan maka tingkat-tingkat kerja berikut ialah simplifikasi, analisa dan syntesis. Untuk dapat memeriksa sesuatu dengan baik dibuatkan dulu suatu bentuk yang ideal (keadaan sempurna) di mana pengaruh-pengaruh dari luar dihilangkan.

Penyusunan bentuk ideal yang disederhanakan disebut *simplifikasi*. Bentuk ideal ini dipecahkan dalam bagian-bagian yang nanti akan dapat lebih mudah diteliti tersendiri; yang disebut *analisa*. Sesudah semua bagian diteliti hasil-hasil dipersatukan lagi, ini disebut *syntesis*.

Sebagai contoh, penelitian tentang pengaruh ledakan bom atom. Kerusakan-kerusakan timbul oleh karena pengaruh ledakan bom atom yang bermacam-macam, yaitu pengaruh sinar, pemindahan udara ke luar, diikuti dengan pemindahan udara ke dalam, getaran tanah, radioactive fallout, dan sebagainya. Ini analisisnya.

Kalau satu komponen saja yang diteliti, umpamanya tekanan udara, maka lebih mudah pengaruh ini diukur dalam satu garis saja, tidak dalam satu bidang apa lagi dalam satu ruang, inilah *simplifikasi*.

Nanti kalau pertanyaan-pertanyaan ini sudah terjawab, diadakan pernyataan sehingga pengaruh di ruang dapat diketahui. Bersama hasil penelitian bagian-bagian lain, akan dilakukan penyatuan yang lebih luas sehingga jawaban pengaruh ledakan atom pada keseluruhannya dapat diberikan.

III. Hipotesa.

Hipotesa ialah hasil pemikiran berupa jawaban pada pertanyaan yang disebut persoalan. Perlu difahami bahwa pada pemikiran ini tak cukup digunakan akal sehat, tetapi diperlukan pemikiran yang logis.

Sebenarnya "common sense" (akal sehat) dan "Science" (ilmu pengetahuan) bertitik tolak dari hal yang sama yaitu,

"There is some order in nature" apa yang terjadi kemarin

juga akan terjadi besok dalam keadaan yang sama.

Perbedaan-perbedaan yang nyata antara dua cara pemikiran ini ialah bahwa ilmu pengetahuan selalu sistematis dan dapat dikontrol.

1. Pada pemikiran akal sehat digunakan teori dan konsep-konsep sesukanya, sedangkan ilmu pengetahuan dengan seksama membangun teori-teorinya dan setiap kali menilai kembali apakah teorinya masih berlaku.

Contoh: Cuaca jelek dan tak teratur disebabkan oleh ledakan bom atom.

2. Seorang ilmuwan menilai teori-teori dan hipotesa-hipotesanya secara sistematis dan empiris.

Contoh: orang Negro pandai musik.

3. Seorang ilmuwan menggunakan kontrol.

Dengan seksama seorang ilmuwan memeriksa apakah yang dilihat tak disebabkan oleh faktor-faktor lain. Dengan lain perkataan, apakah tak ada lain variabel memegang peranan kecuali yang disebut dalam hipotesa.

4. Seorang ilmuwan selalu mencari hubungan antara dua kejadian (relation among phenomena). Hal ini juga dilakukan orang awam mudah sekali menghubungkan dua kejadian yang terjadi bersama-sama, tanpa merasa perlu membuktikan hubungan ini.

5. Seorang ilmuwan dengan teliti menyingkirkan apa yang disebut "Metaphysical Phenomena". Yang dimaksud metaphysical phenomena ialah suatu hal yang tak dapat ditest.

"Science is concerned with things that can be publicly observed and tested".

Hipotesa ialah jawaban pada pertanyaan yang disebut persoalan maka hipotesa ialah suatu pertanyaan yang menghubungkan dua variabel sedemikian rupa sehingga hubungan ini dapat dibuktikan secara empiris.

Dalam mencari hipotesa sebenarnya dicari penyebab, umpamanya:

(a) *Method of agreement.*

Penyebab ialah faktor yang terdapat pada semua keadaan di mana persoalan yang diselidiki itu timbul.

Kalau umpamanya terdapat rentetan pembunuhan di suatu kota dan pada setiap pembunuhan itu dilihat oleh saksi-saksi bahwa di dekat tempat pembunuhan itu selalu ada seorang tertentu, maka polisi akan sangat curiga bahwa orang ini ialah si pembunuh.

Kalau suatu penyakit hanya terdapat pada orang - orang yang hidup atau pernah hidup di negara tropis, maka tak jauh dari benar anggapan bahwa penyakit ini ialah penyakit tropis.

(b) *Method of difference.*

Kalau dua hal hanya berbeda dalam satu segi saja, kemudian timbul suatu perbedaan, maka segi tersebut sangat mungkin penyebabnya.

(c) *Principal of concomitant variation.*

Kalau variasi intensitas suatu faktor berjalan sejajar dengan efek intensitas, maka faktor tersebut dapat dianggap sebagai penyebab. Sebagai contoh seorang membuat limun dan terdapat keluhan bahwa limun terasa pahit.

Kalau rasa pahit berjalan sejajar dengan warna merah limun, atau dengan lain perkataan makin merah limun makin pahit rasanya, maka mungkin sekali bahan pemberi warna ialah penyebab rasa pahit itu.

Masih dua hal yang perlu disebut untuk melengkapi gambaran umum ini yaitu sedikit pengertian tentang sampling dan kegunaan ilmu statistik.

Jika kita menghendaki informasi tertentu (data) dari suatu kelompok (universe, population), informasi ini

dapat diambil dengan menghubungkan setiap unsur kelompok satu persatu. Tampaknya seakan-akan dengan cara ini akan terkumpul informasi yang paling lengkap. Dalam kenyataan hal ini tidak benar. Tiap penelitian terikat pada jumlah tenaga, waktu, dan keuangan.

Pada kelompok besar, faktor-faktor tersebut di atas akan tidak memungkinkan menghubungi setiap unsur satu persatu. Jika dibutuhkan banyak tenaga, akibatnya ialah bahwa akan digunakan tenaga-tenaga yang kurang pengalaman dan yang juga tak mengerti maksud tujuan penelitian.

Oleh karena itu akan terkumpul nanti jawaban-jawaban yang banyak bias-nya. Kalau data dikumpulkan dari seluruh kelompok maka akan terkumpul begitu banyak data sehingga mengolahnya (processing) tak dapat dilaksanakan dengan baik dan tertib. Untuk mengatasi persoalan ini maka digunakan teknik sampling. Sample ialah sebagian dari kelompok (universe, population) yang akan diteliti dan yang masih mempunyai semua ciri-ciri kelompok asalnya. Tentu saja mengambil sample ini ada cara-cara tertentu yang akan dibicarakan tersendiri.

Untuk menentukan sample, menyusun, dan mengolah data, dan mengambil kesimpulan diperlukan perhitungan - perhitungan. Untuk perhitungan inilah diperlukan ilmu statistik. Maksud utama statistik untuk penelitian ialah memperoleh informasi sebanyak mungkin dari data sedikit mungkin, dengan mengetahui berapa besar kemungkinan kesimpulan itu betul atau salah.

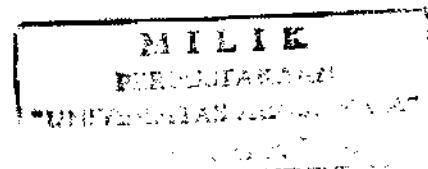
Ilmu statistik ini akan dibicarakan tersendiri dengan memilih bahan yang diperlukan untuk penelitian.

Dengan demikian telah kami coba menyampaikan garis-garis besar metodik penelitian dengan maksud terdapat suatu gambaran umum dan dasar untuk pembicaraan selanjutnya.

1

BUKU-BUKU YANG DIGUNAKAN SEBAGAI SUMBER KEPUSTAKAAN

1. ACKOFF, R.L.,
Scientific Method, John Wiley & Sons, Inc. (1962)
2. BRADFORD HILL, A.,
Principles of Medical Statistics, Oxford University Press (1966)^{8th} edition.
3. BRIGHT WILSON JR., E.,
An Introduction to Scientific Research, Mc. Graw Hill Book Company Inc.
4. BUCKLEY, W.,
Modern Systems Research for the Behavioral Scientist. Aldine Publishing Company, Chicago (1968).
5. *Encyclopaedia Britannica*.
Encyclopaedia Britannica Inc. William Benton Publisher (1968).
6. KERLINGER, F.N.,
Foundations of Behavioral Research, Holt, Rinehart & Winston.



2. STATISTIK DESKRIPTIF

Di sini akan dibicarakan cara penyusunan data yang dikumpulkan supaya menjadi lebih jelas, serta perhitungan angka-angka yang dapat menerangkan beberapa macam sifat data itu.

Catatan: Dalam bidang ilmu ini, istilah "statistik" berarti suatu angka yang merupakan suatu ukuran alam semesta.

Jadi data terdiri dari suatu kumpulan statistik.

Dalam teori kesimpulan statistik, istilah "statistik" berarti lain.

2.1. Tabel Statistik.

Tabel Statistik adalah susunan sistematis dari data angka yang disajikan dalam kolom dan baris untuk maksud perbandingan.

Tabel Umum ialah yang menyajikan data asli.

Tabel Khusus adalah merupakan suatu penyederhanaan tabel umum yang dipergunakan untuk mengutamakan hubungan-hubungan tertentu. Seringkali bilangan-bilangan pembulatan yang dipergunakan.

Ada aturan-aturan tentang penyusunan tabel serta penyusunan data di dalam tabel yang telah diterima oleh umum, yang seharusnya diperhatikan pada waktu menyusun suatu tabel yang akan diterbitkan nanti.

2.2. Grafik.

Grafik adalah cara penyajian data dalam bentuk gambar. Ada juga aturan-aturan yang perlu diperhatikan dalam penyusunan grafik. Ada beberapa macam grafik.

a. *Grafik garis.*

Ini adalah macam grafik yang paling umum, yang dibentuk dengan menghubungkan titik-titik yang merupakan data dengan suatu garis.

Biasanya suatu skala aritmetik yang dipergunakan, dan dalam hal ini titik nol harus ditunjukkan pada skala.

Kalau skalanya dalam bentuk persentase maka garis 100% juga harus ditunjukkan.

Apabila diinginkan untuk membandingkan perubahan-perubahan relatif dalam suatu variabel, maka seharusnya dipergunakan suatu skala logaritmik untuk data variabel itu. Kalau skala logaritmik yang dipergunakan pada kedua sumbu, maka skala grafik itu disebut logaritmik. Kalau skala logaritmik dipergunakan pada satu sumbu, maka skala grafik itu disebut semi logaritmik.

b. Grafik garis khusus

Termasuk dalam grafik garis khusus ini ialah silhoutte chart, band chart, high low graph dan histogram.

Hanya histogram saja yang banyak dipergunakan, yang cara penyusunannya sbb. :

Empat persegi panjang digambar dengan alasnya lebar interval - interval kelas dan tingginya frekuensi interval-interval itu.

Contoh histogram :

Tabel

Distribusi frekuensi hasil ujian statistik deskriptif oleh 111 Mahasiswa persiapan FEUI, 1967.

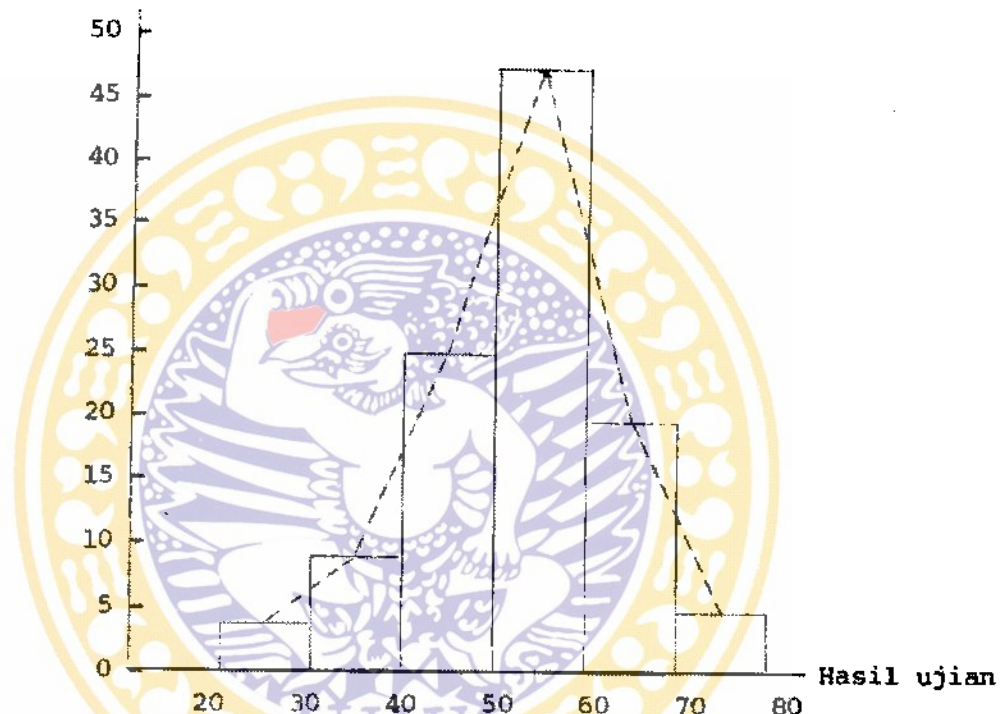
Hasil Ujian	Jumlah Mahasiswa
20 - 29,99	4
30 - 39,99	9
40 - 49,99	25
50 - 59,99	48
60 - 69,99	20
70 - 79,99	5

Sumber : Biro Pendidikan Fakultas Ekonomi,
Universitas Indonesia, 1967.

Diagram

Hasil ujian statistik deskriptif III mahasiswa tingkat persiapan FEUI 1967, yang digambarkan dengan diagram kolom serta kurva frekuensi.

Jumlah mahasiswa



c. Bar Chart, (Peta batang)

Bar chart menggambarkan perbedaan angka dengan membandingkan batang-batang dengan tinggi yang berbeda-beda tetapi dengan lebar yang sama.

Dapat berbentuk sederhana atau terbagi-bagi, dengan skala absolut atau skala persentase.

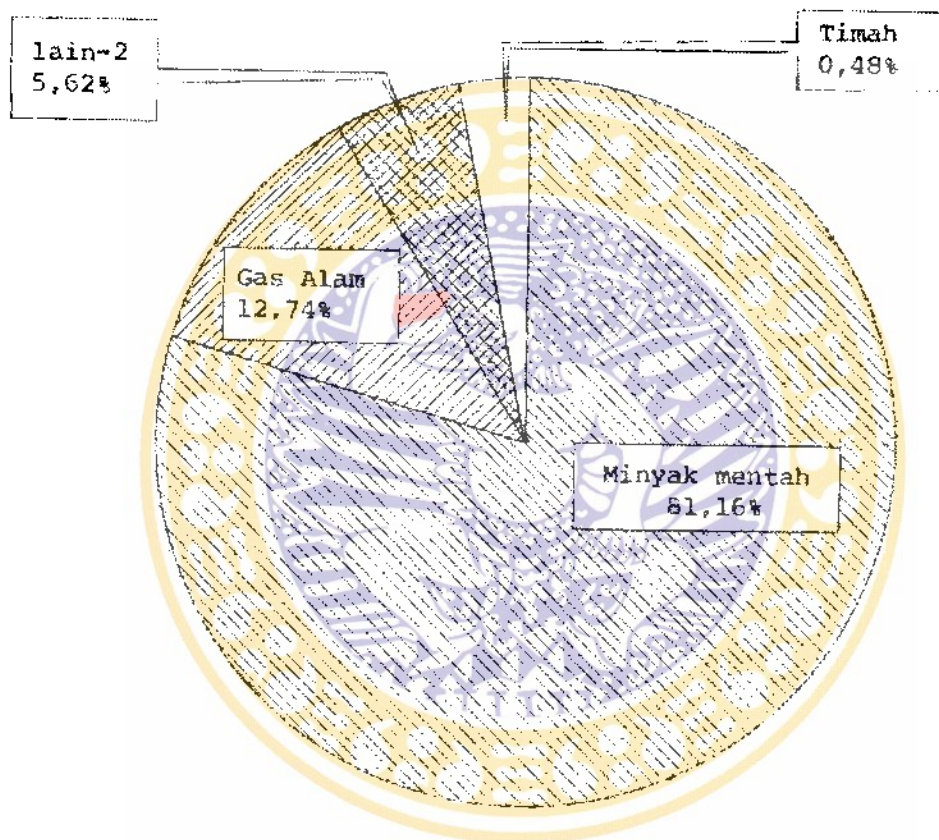
Peta batang dapat disusun secara vertikal maupun horisontal.

Bila data dapat diklasifikasikan secara kronologis, maka peta batangnya sebaiknya disusun secara vertikal.

Diagram di bawah ini melukiskan penyusunan peta batang secara vertikal.

Pie chart (diagram lingkaran).

Diagram
Produksi tambang di Indonesia
1963



Keterangan:

Minyak mentah	: 81,16%
Gas alam	: 12,74%
Lain-lain	: 5,62%
Timah	: 0,48%

Sumber : Indonesia, Facts and Figures. Hal 356, oleh Nugroho.

e. Solid Diagram

Solid diagram menunjukkan perbedaan angka-angka dengan isi yang berbeda-beda misalnya isi kubus, silinder, dlsb.

f. Map Graph.

2.3. Distribusi Frekuensi.

Distribusi frekuensi adalah susunan data angka menurut besarnya (distribusi frekuensi kuantitatif) atau golongannya (distribusi frekuensi kualitatif).

Contoh-contoh: 1. Distribusi frekuensi kuantitatif : lihat bagian berikut.

2. Distribusi frekuensi kualitatif : hasil suatu survey untuk menyelidiki hubungan antara merokok dengan mati akibat sakit kanker adalah sbb. :

Tidak merokok dan tidak mati sakit kanker : 453 orang
 Tidak merokok dan mati sakit kanker : 28 orang
 Merokok dan tidak mati sakit kanker : 581 orang
 Merokok dan mati sakit kanker : 52 orang

	Tidak mati sakit kanker	mati sakit kanker
Not smoking	453	28
Smoking	581	52

Di sini terdapat 2 cara penggolongan. Data ini bila disajikan dalam bentuk tabel disebut "Contingency Table".

Jumlah cara penggolongan dapat 1,2,3,, tetapi lebih dari tiga cara penggolongan hampir tidak pernah dipakai.

Penyusunan Distribusi Frekuensi Kuantitatif.

Langkah pertama ialah pemilihan banyaknya interval klas. Biasanya distribusi yang memakai 5 - 15 interval klas yang paling bermanfaat, tergantung pada banyaknya dan penyebaran data.

Suatu rumus yang dianjurkan untuk menentukan banyaknya interval kelas, K ialah:

$$K = 1 + 3,3 \log n,$$

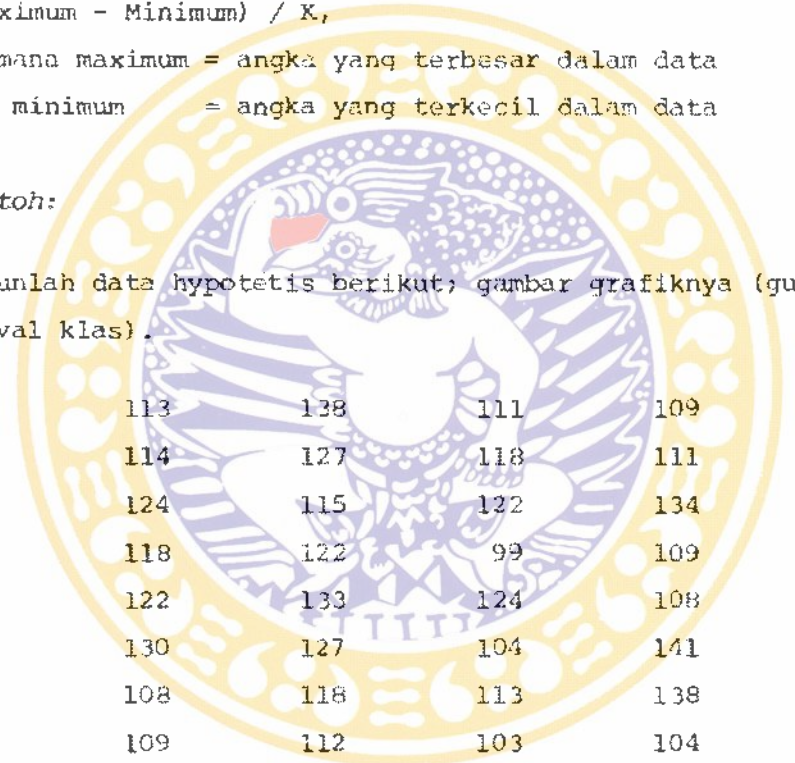
di mana n = banyaknya data.

Kemudian ditentukan lebar interval kelas menurut lebar interval = (Maximum - Minimum) / K ,

di mana maximum = angka yang terbesar dalam data
dan minimum = angka yang terkecil dalam data

Contoh:

Susunlah data hipotetis berikut; gambar grafiknya (gunakan 10 interval kelas).



102	113	138	111	109
99	114	127	118	111
130	124	115	122	134
108	118	122	99	109
106	122	133	124	108
102	130	127	104	141
103	108	118	113	138
101	109	112	103	104
111	106	126	114	102
107	146	108	109	114

Penyelesaian

$K = 10$ buah

$$\text{Lebar interval} = \frac{146 - 99}{10} = \frac{47}{10} = 4,7$$

Lebar interval dapat diambil 4 atau 5

Andaikata kita ambil 5, maka terdapat distribusi frekuensi sebagai berikut:

Interval Klas	Tabulasi	Frekuensi f_i	Frekuensi Kumulatif	Mid Point (x_i)
98 - 102	11111 1	6		100
103 - 107	11111 11	7	13	105
108 - 112	11111 11111 11	12	25	110
113 - 117	11111 1	6	31	115
118 - 122	11111 1	6	37	120
123 - 127	11111	5	42	125
128 - 132	11	2	44	130
133 - 137	11	2	46	135
138 - 142	111	3	49	140
143 - 147	1	1	50	145
		$\Sigma f_i = 50$		

Perhatikanlah bahwa angka-angka yang ditunjukkan di dalam kolom pertama tabel di atas adalah yang disebutkan "*class limits*", yang diperlukan nanti dalam beberapa perhitungan dengan "*grouped data*" seperti ini ialah "*class boundaries*", yang dalam contoh ini adalah : 97,5 ; 102,5 ; 107,5 ; "*class width*" / lebar interval = jarak antara "*class boundaries*", = 5 dalam contoh ini.

Distribusi frekuensi ini dapat dipergunakan untuk membentuk suatu grafik garis histogram, grafik garis polygon, dlsb.

2.4. Harga Menengah.

a. Rata-rata hitung

Andaikan bahwa data terdiri dari bilangan-bilangan x_1, x_2, \dots, x_n , maka rata-rata hitung \bar{x} yang juga disebut "*sample mean*", diberikan oleh rumus:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Secara lebih umum, andaikan bahwa bilangan x_i terdapat sebanyak f_i kali dalam data ($i = 1, 2, \dots, k$ di mana $k =$ jumlah bilangan yang berbeda-beda dalam data), maka

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{i=1}^k f_i}$$

Perhatikanlah bahwa:

$$\sum_{i=1}^k f_i = n$$

Catatan 1 : Bentuk yang terakhir ini, dengan "mid-point" klas sebagai x_i ($i = 1, 2, \dots, k$), yang diperlukan apabila data disajikan dalam bentuk distribusi frekuensi.

Catatan 2 : Kadang-kadang perhitungan dapat dipermudah dengan memakai suatu skala baru.

b. Median

Median sekumpulan angka adalah harga yang di tengah bila angka itu disusun menurut besarnya. Kalau sekumpulan angka itu banyaknya genap, maka median adalah rata-rata hitung dari dua harga yang di tengah.

Apabila datanya dikelompokkan dalam beberapa interval klas, harga median dapat dicari dengan rumus:

$$Med = l_{Me} + i (n/2 - \sum_1 f) / f_{Me}$$

di mana l_{Me} = batas bawah ("lower class boundary") interval yang memuat median.

n = banyaknya data

$\sum_1 f$ = jumlah semua frekuensi dalam interval-interval sebelum interval yang memuat median.

f_{Me} = frekuensi interval yang memuat median.

i = lebar interval.

c. Mode

Mode adalah harga yang frekuensinya paling besar. Untuk data yang dikelompokkan dalam beberapa interval klas, harga mode dapat dicari dengan rumus :

$$M_o = l_{M_o} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot i$$

dimana l_{M_o} = batas bawah interval mode.

Δ_1 = beda antara frekuensi interval mode dengan interval sebelumnya.

Δ_2 = beda antara frekuensi interval mode dengan interval sesudahnya.

i = lebar interval.

Interval mode adalah interval yang mempunyai frekuensi tertinggi.

Catatan 1 : Berhubungan dengan data yang dikelompokkan dalam suatu distribusi frekuensi, perlu disebutkan bahwa harga yang terdapat untuk mean, median dan mode dipengaruhi oleh cara penyusunan distribusi frekuensi itu (pemilihan lebar interval dsb.), apalagi bila data tidak banyak.

Catatan 2 : Mode dan median tidak dipengaruhi oleh harga-harga yang luar biasa, jadi merekalah yang sering dipergunakan sebagai harga menengah bila ada harga-harga yang luar biasa.

Bila sample besar sehingga letak mode cukup jelas, maka mode yang biasanya dipergunakan.

d. *Geometric mean*

Bila dipunyai sekumpulan angka-angka x_1, x_2, \dots, x_n maka geometric mean didefinisikan sbb. :

$$GM = \sqrt[n]{(x_1, x_2, \dots, x_n)}$$

Geometric mean sering dipergunakan dengan data dalam bentuk persentase-persentase perubahan.

2. Harga-harga Penyimpangan.

Harga tengah kurang berguna bila tidak diketahui pemencaran atau penyimpangan tiap datanya terhadap harga tengah tadi. Bila harga penyimpangan tiap datanya terhadap harga tengah sangat besar, maka harga tengah tadi kurang dapat menggambarkan keadaan datanya.

Ukuran terpencarnya data:

- (1) Range
- (2) Penyimpangan rata-rata (average deviation)
- (3) Penyimpangan standard (standard deviation)
- (4) Penyimpangan kuartil (quartile deviation)

(1) *Range:*

Range adalah harga penyimpangan yang paling sederhana, didefinisikan sebagai perbedaan harga yang tertinggi dan harga terendah. Range memberi gambaran seberapa jauh data memencar.

Contoh: Tekanan darah systolic 12 pemuda adalah 116, 130, 126, 110, 114, 142, 140, 132, 120, 112, 118, 118 (dalam mm Hg).

Harga yang tertinggi = 142 mm Hg.

Harga yang terendah = 110 mm Hg.

Range = 142 - 110 = 32 mm Hg.

(2) Penyimpangan rata-rata (Mean Deviation)

Penyimpangan rata-rata adalah harga rata-rata dari penyimpangan-penyimpangan tiap datanya terhadap meannya. Makin kecil harga penyimpangan berarti makin kecil pemencaran data tadi terhadap harga meannya.

Penyimpangan rata-rata dihitung dengan rumus:

$$MD = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n}$$

Contoh: Tekanan darah systolic 12 orang pemuda adalah

Tekanan darah (x_i)	$x_i - \bar{x}$
116	- 7,2
130	+ 6,8
126	+ 2,8
110	- 13,2
114	- 9,2
142	+ 18,8
140	+ 16,8
132	+ 8,8
120	- 3,2
112	- 11,2
118	- 5,2
118	- 5,2
$\sum x_i = 1.478$	108,4

$$\bar{x} = 123,166 = 123,2 \text{ mm Hg.}$$

$$MD = \frac{108,4}{12} = 9,033 \text{ mm Hg.}$$

Bila datanya disajikan dalam bentuk interval klas maka rumus mean deviation ialah:

$$MD = \frac{\sum_{i=1}^k f_i \cdot (x_i - \bar{x})}{\sum_{i=1}^k f_i}$$

(3) Penyimpangan standard (Standard Deviation)

Definisi:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

(s^2 yang disebut "sample variance").

Bila datanya disajikan dalam interval klas, maka bentuk definisi ini ialah:

$$s = \sqrt{\frac{\sum_{i=1}^k f_i (x_i - \bar{x})^2}{\left(\sum_{i=1}^k f_i\right) - 1}}$$

$$\sum_{i=1}^k f_i = n = \text{banyaknya data}$$

Untuk perhitungan dengan memakai mesin hitung bentuk lain yang paling berguna ialah:

$$s^2 = \frac{1}{n - 1} \left\{ \sum_{i=1}^k f_i x_i^2 - \frac{\left(\sum_{i=1}^k f_i x_i \right)^2}{n} \right\}$$

atau

$$s^2 = \frac{1}{n - 1} \left\{ \sum_{i=1}^k f_i x_i^2 - n \bar{x}^2 \right\}$$

Untuk perhitungan s^2 dari distribusi frekuensi dengan memakai-tangan, seharusnya kita memakai bentuk lain lagi yang berlaku bila skala baru dipakai, ialah:

$$s = i \sqrt{\frac{n \sum_{i=1}^k f_i d_i^2 - (\sum_{i=1}^k f_i d_i)^2}{n(n-1)}}$$

dimana:

- i = lebar interval
- f_i = frekuensi interval ke - i
- d_i = skala baru yang dipakai
- n = banyaknya data
- k = banyaknya interval.

Contoh: data (hipotetis)

Mid point x_i	f_i	Skala baru d_i	d_i^2	$f_i d_i$	$f_i d_i^2$
100	6	- 4	16	- 24	96
105	7	- 3	9	- 21	63
110	12	- 2	4	- 24	48
115	6	- 1	1	- 6	6
120	6	0	0	0	0
125	5	1	1	5	5
130	2	2	4	4	8
135	2	3	9	6	18
140	3	4	16	12	48
145	1	5	25	5	25
Total	50	-	-	- 43	317

$$\begin{aligned}
 i &= 5 \\
 s &= 5 \sqrt{\frac{50(317) - (-43)^2}{50.49}} = 5 \sqrt{\frac{15850 - 1849}{2450}} \\
 &= 5 \sqrt{\frac{14001}{2450}} = 5 \sqrt{5,714} = 11,952
 \end{aligned}$$

(4) Penyimpangan kuartil

Definisi:
$$Q = \frac{Q_3 - Q_1}{2}$$

dimana Q_1 (kuartil pertama) adalah harga yang sedemikian besar sehingga seperempat data kurang dari Q_1 , sedangkan tigaperempat data kurang dari Q_3 .

$$Q_1 = l_{Q_1} + \frac{(\frac{n}{4} - \Sigma, f)}{f_{Q_1}}$$

l_{Q_1} = batas bawah interval yang memuat Q_1

n = banyaknya data

Σ, f = jumlah semua frekuensi interval-interval sebelum interval yang memuat Q_1

f_{Q_1} = frekuensi interval yang memuat Q_1

i = lebar interval.

$$Q_3 = l_{Q_3} + \frac{i (\frac{3n}{4} - \Sigma, f)}{f_{Q_3}}$$

Harga-harga kuartil membagi data menjadi empat bagian yang sama besarnya. Perhitungan-perhitungannya seperti perhitungan median, dan Q_2 = median.

Dua Harga Penyimpangan Relatif

Koefisien variasi Pearson: $V = \frac{s}{x} \cdot 100$

Penyimpangan kuartil relatif : $V_{pq} = (Q_3 - Q_1) / (Q_3 + Q_1)$

2.6. Kemencengan (Skewness)

Harga kemencengan adalah harga yang menunjukkan seberapa jauh distribusi itu menyimpang dari simetris. Apabila suatu distribusi itu simetris, maka harga-harga mean, median, dan mode berimpit (sama besar). Bila tidak simetris harga-harga tadi tidak sama besar.

Karena harga mean cenderung untuk bergeser ke arah harga-harga yang terlalu besar (kecil) sedangkan mode tidak dipengaruhi oleh harga-harga semacam itu, maka makin menceng distribusinya makin besar jarak antara mean dengan mode. Oleh karena itu jarak antara mean dengan mode ini dapat digunakan sebagai ukuran kemencengan distribusinya.

Karena biasanya harga kemencengan ini digunakan untuk membandingkan kemencengan dua distribusi atau lebih maka untuk menghindari kemungkinan adanya satuan yang berbeda dari distribusi-distribusi itu dan juga untuk menghindari pengaruh harga penyimpangan dalam perbandingan ini didefinisikan harga kemencengan KM sebagai berikut:

$$KM = \frac{\text{mean} - \text{mode}}{\text{penyimpangan standard}}$$

Untuk distribusi yang tidak terlalu menceng rumus di atas dapat diganti dengan:

$$KM = \frac{3 (\text{mean} - \text{median})}{\text{penyimpangan standard}}$$

Dari rumus definisi ini jelas bahwa untuk distribusi simetris harga kemencengannya sama dengan nol.

Untuk distribusi yang mempunyai mean lebih besar dari mode, harga kemencengannya positif dan distribusinya dinamakan menceng positif (ke kanan).

Sebaliknya bila mean lebih kecil dari mode, harga kemencengannya negatif dan distribusinya dinamakan menceng negatif (ke kiri).



Harga kemencengan dapat pula diukur dengan melihat jarak antara harga-harga kuartil. Pada distribusi simetris jarak antara kuartil pertama dengan median sama dengan jarak antara median dengan kuartil ketiga. Pada distribusi yang menceng (tidak simetris) jarak-jarak tadi tidak sama.

Dengan mengingat kenyataan ini, suatu harga kemencengan dapat didefinisikan sebagai berikut:

$$KM = \frac{(Q_3 - \text{Median}) - (\text{Median} - Q_1)}{\text{penyimpangan kuartil}}$$

Dari definisi terlihat bahwa untuk distribusi simetris, $KM = 0$. Bila $KM > 0$ distribusi menceng ke kanan dan bila $KM < 0$ distribusi menceng ke kiri.

APPENDIX

tentang

ATURAN-ATURAN PENYUSUNAN TABEL

Cara-cara penyusunan tabel yang biasa dilakukan adalah bermacam-macam. Namun demikian aturan-aturan yang umumnya diikuti dalam penyusunan tabel statistik adalah sebagai berikut.

1. *Titel*

Titel harus jelas dan singkat dan harus pula menunjukkan hal-hal yang berikut:

- a. Keadaan data yang disajikan
- b. Daerah yang meliputinya
- c. Jangka periode waktunya.

Titel diletakkan di atas tabel. Huruf yang dipergunakan untuk judul lebih besar dari bagian-bagian lain.

2. *Sumber*

Sumber data harus ditunjukkan dalam tabel. Kecuali untuk menunjukkan dari mana data aslinya didapat juga untuk:

- a. Menunjukkan yang memiliki data itu
- b. Maksud pemeriksaan
- c. Reference kalau menambah data.

3. *Catatan (Footnote)*

Catatan digunakan untuk memberi keterangan lebih lanjut tentang angka-angka dalam tabel, dst.

Catatan diletakkan antara tabel dan sumber.

Catatan selalu ditunjukkan dengan lambang, misalnya +, #, * dst atau dengan huruf-huruf, jangan digunakan angka, sebab dapat kacau dengan angka-angka dalam tabel.

3. STATISTIK ANALITIS

Random Variabel dan Distribusi Probabilitasnya.

Teori statistik analitis (kesimpulan statistik) didasarkan atas teori probabilitas, karena itu untuk dapat memahami cara bekerja statistik analitis - pada khususnya supaya dapat memahami pengertian "level of significance" dalam pengujian hipotesa - perlu kita membahas lebih dahulu beberapa pengertian pokok teori probabilitas itu.

Definisi-definisi:

Random variabel adalah bilangan nyata fungsi yang didapat dari hasil suatu percobaan.

Penyebaran probabilitas di antara semua harga yang dicapai oleh suatu random variabel (r.v.) disebut distribusi probabilitasnya.

Bagi r.v. yang kontinu - artinya harga-harganya terdiri dari semua bilangan nyata di antara batas-batas tertentu, distribusi probabilitas itu dinyatakan dengan melalui fungsi densitas (kepadatan) probabilitas $f_x(X)$ atau integralnya yaitu fungsi distribusi $F_x(X)$

Pr. (x terdapat di dalam $f_x(x+dx) = f_x(x) dx$
Df

$F_x(x) = \int_{-\infty}^x f_x(t) dt = \text{Pr. (x terdapat di dalam } -\infty, x).$

Fungsi distribusilah yang ditabelkan buat distribusi-distribusi yang penting.

Beberapa kesimpulan mengenai fungsi distribusi:

- (i) $F_x(-\infty) = 0$ sedangkan $F_x(\infty) = 1$
- (ii) $F_x(b) - F_x(a) = \int_a^b f_x(t) dt - \int_{-\infty}^a f_x(t) dt = \int_a^b f_x(t) dt$
(b > a)
 $= \text{Pr } < x \text{ terdapat di dalam } (a, b) >$

Contoh-contoh:

1. Distribusi normal

Parameter-parameternya dua yang biasanya disebut μ dan σ^2 dan distribusinya ditulis secara singkat $N(\mu, \sigma^2)$.

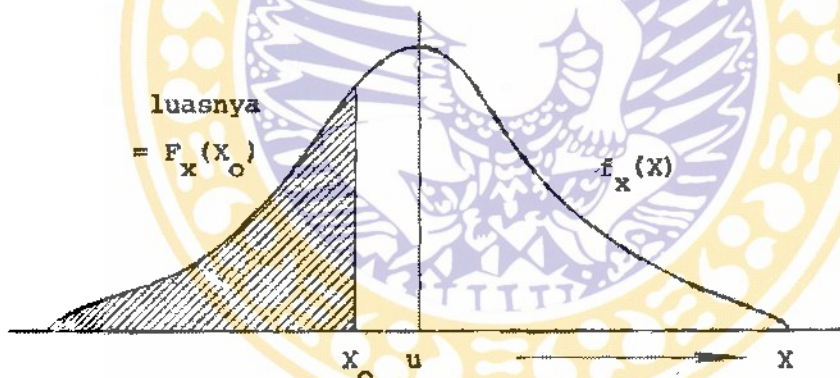
Selanjutnya, kalau $x \sim N(\mu, \sigma^2)$ berarti r.v.x. mempunyai distribusi $N(\mu, \sigma^2)$.

Maka $y \stackrel{\text{Df}}{=} \frac{x-\mu}{\sigma} \sim N(0, 1)$ distribusi normal standard.

Hanya distribusi $N(0, 1)$ yang ditabelkan karena dengan memakai tabelnya semua probabilitas yang bersangkutan dengan distribusi normal dapat dihitung:

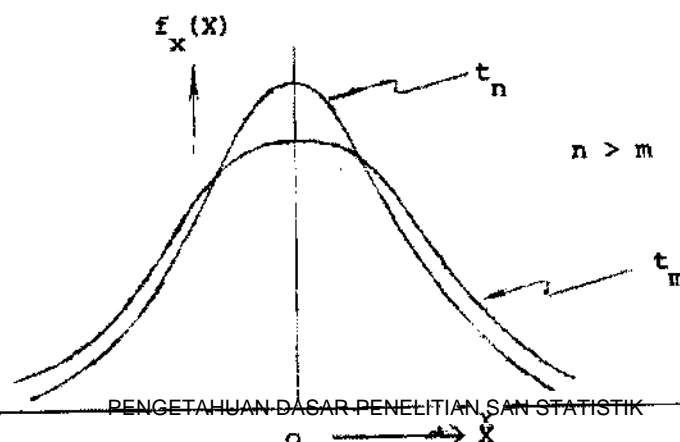
$$\begin{aligned} \Pr(a < x < b) &= \Pr\left(\frac{a-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) \\ &= \Pr\left(\frac{a-\mu}{\sigma} < y < \frac{b-\mu}{\sigma}\right) \end{aligned}$$

Perhatikanlah sifat simetris distribusi normal.



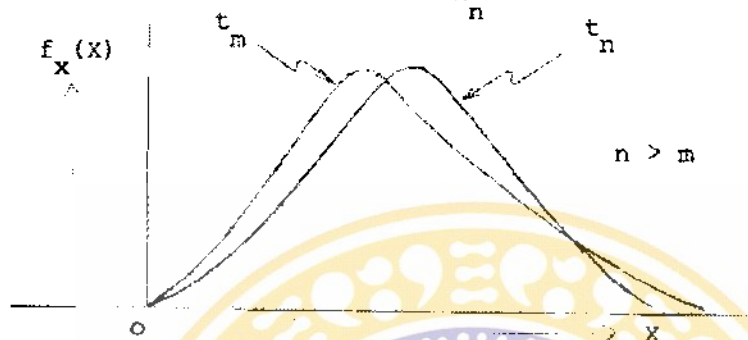
2. Distribusi t (student)

Parameternya satu yang disebut derajat kebebasan (degrees of freedom = d.f.) dan distribusinya ditulis secara singkat t_n (n selalu bilangan bulat). t_n mendekati $N(0, 1)$ bila $n \rightarrow \infty$, dan pendekatan biasanya dianggap cukup baik apabila $n \geq 30$.



3. Distribusi Kai Kwadrat, (Chi Square)

Parameternya satu yang disebut derajat kebebasan dan distribusinya ditulis secara singkat χ^2_n



4. Distribusi F (Variance Ratio)

Harga-harganya terdiri dari bilangan-bilangan positif seperti halnya χ^2_n tersebut di atas.

Parameternya dua, yang juga disebut derajat kebebasan dan distribusinya ditulis secara singkat $F_{n,m}$.

Statistik.

Definisi: Statistik adalah r.v. fungsi dari data yang dikumpulkan dari suatu percobaan/survey.

Contoh: Misalnya data tertentu dari x_1, x_2, \dots, x_n , maka $x_1 + x_n, x_1 / x_2$,

$$\text{sample mean } \bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

$$\text{sample variance } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

merupakan statistik-statistik.

Statistik-statistik dipakai dalam penaksiran parameter-parameter serta pengujian hipotesa. Untuk tujuan itu distribusi probabilitasnya harus dihitung lebih dahulu didasarkan atas pengandaian-pengandaian mengenai data dan dengan memakai teori probabilitas.

Pengujian Hipotesa.

Hipotesa (dugaan) pokok yang kebenarannya harus diselidiki diberi nama hipotesa nol (H_0).

Misalnya, apabila distribusi normal yang ditinjau, contoh hipotesa nol ialah $H_0: u = u_0$

Hipotesa Alternatif (lawan) H_1 ialah hipotesa yang dianggap dapat berlaku dalam hal hipotesa nol tidak tepat.

Hipotesa ini dikatakan "one-sided" (searah) hanya kalau memuat harga-harga yang lebih besar dari pada yang tercantum di dalam H_0 , atau hanya harga-harga yang lebih kecil, misalnya $H_1: u > u_0$ atau $H_1: u < u_0$. Kalau baik harga-harga yang lebih besar maupun harga-harga yang lebih kecil yang tercantum, maka hipotesa lawan itu disebut "two sided" (dwi arah) misalnya, $u \neq u_0$.

Bagi setiap cara pengujian hipotesa, probabilitas menolak hipotesa apabila hipotesa itu memang benar yang disebut "level of significance" (size of test) dari pada cara (test) yang bersangkutan.

Sebenarnya ada dua macam kesalahan yang dapat timbul dalam pengujian hipotesa, untuk jelasnya lihat gambar di bawah ini:

	H_0 benar	H_1 benar
H_0 diterima	Keputusan betul	Keputusan salah (probabilitasnya α)
H_0 ditolak	Keputusan salah (probabilitasnya α)	Keputusan betul

α yang disebut "power of test" jadi power adalah probability menolak hipotesa nol yang tidak benar.

Pengujian hipotesa tentang mean distribusi normal dan tentang selisih antara kedua mean dua distribusi normal.

(i) Soal: Diketahui $x \sim N(u, \sigma^2)$ di mana σ^2 tidak diketahui.

Disajikan data bilangan-bilangan nyata x_1, x_2, \dots, x_n

H_0 : mean $u = u_0$

Penyelesaian:

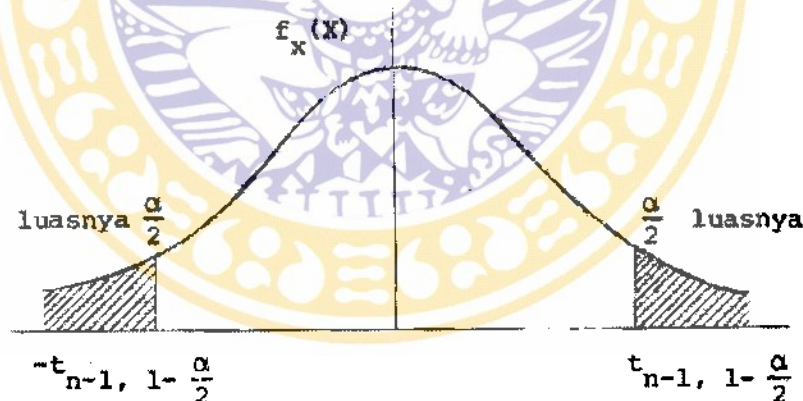
Diketahui bahwa $\frac{\bar{x} - u_0}{s / \sqrt{n}} \sim t_{(n-1)}$ bila H_0 benar

(Mengenai definisi \bar{x} dan s lihat fasal 6)

$$\text{Hitung } t = \frac{\bar{x} - u_0}{s / \sqrt{n}}$$

Bila H_1 adalah " $u \neq u_0$ " maka H_0 harus ditolak apabila

$|t| > t_{(n-1); (1-\alpha/2)}$ supaya level of significance α dapat dicapai, untuk jelasnya lihat gambar:



Perhatikan bahwa $t_{n,p}$ didefinisikan dengan melalui rumus

Rumus $F_y(t_{n,p}) = p$ dimana $y \sim t_n$

Bila H_1 adalah " $u < u_0$ " maka H_0 harus ditolak apabila $t < -t_{(n-1); (1-\alpha/2)}$

Bila H_1 adalah " $u > u_0$ " maka H_0 harus ditolak apabila $t > t_{(n-1); (1-\alpha/2)}$

Harga-harga $t_{(n-1); (1-\alpha/2)}$ dan $t_{(n-1); (1-\alpha/2)}$ yang didapatkan di dalam tabel.

Contoh :

Menurut si penjual, mesin stensil "Roneo" dapat menstensi-
sil sejam 6.500 helai kertas. Sebuah perusahaan persten-
silah ingin membuktikan kebenaran hipotesa itu. Perusa-
haan mengadakan peninjauan dengan menggunakan 12 buah
mesin "Roneo" dan hasil pengamatan sedemikian itu dibe-
rikan seperti berikut,

6.000	5.900	6.200	5.500	6.100	6.200
5.800	6.400	6.500	5.400	6.200	6.700

Apakah ada alasan bagi perusahaan guna menolak hipotesa
tersebut?

Penyelesaian : $H_0 : u = 6.500$

$H_1 : u \neq 6.500$

Andaikan distribusi banyaknya helai sejam normal menurut
hasil di atas, maka H_0 harus ditolak apabila:

$$\left| \frac{\bar{x} - 6.500}{s/\sqrt{12}} \right| > t_{11; 0,975}$$

(Kita pakai $\alpha = 0,05$ bila harganya tidak disebutkan).

Disini $\bar{x} = 6.075$ dan $s = 384,06$

Sehingga $\frac{\bar{x} - 6.500}{s/\sqrt{12}} = - 3,818$, sedangkan $t_{11; 0,975} = 2,201$

Jadi H_0 harus ditolak.

(ii) Soal: Diketahui $x_1 \sim N(u_1, \sigma^2)$ dan $x_2 \sim N(u_2, \sigma^2)$,

dimana σ^2 tidak diketahui. Disajikan data bilangan-

bilangan nyata $x_{11}, x_{12}, \dots, x_{1n_1}$ berupa random

sample dari r.v. x_1 , serta bilangan-bilangan nyata $x_{21},$

x_{22}, \dots, x_{2n_2} berupa random sample dari r.v. x_2 , di ma-

na tidak ada hubungan antara pemilihan sample pertama
dan sample kedua (unpaired data).

$$H_0 : u_1 = u_2$$

Penyelesaian:

Diketahui bahwa
$$\frac{\bar{x}_1 - \bar{x}_2}{\text{Sp} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1 + n_2 - 2)}$$

bila H_0 benar dimana
$$\sigma_P^2 = \frac{(n_1 - 1) \delta_1^2 + (n_2 - 1) \delta_2^2}{n_1 + n_2 - 2}$$

∴ Hitung t .

Bila H_1 adalah " $\mu_1 \neq \mu_2$ " maka H_0 harus ditolak apabila $|t| > t_{(n_1 + n_2 - 2)} \cdot (1 - \alpha/2)$

Bila H_1 adalah " $\mu_1 > \mu_2$ " maka H_0 harus ditolak apabila $t > t_{(n_1 + n_2 - 2)} \cdot (1 - \alpha/2)$

Bila H_1 adalah " $\mu_1 < \mu_2$ " maka H_0 harus ditolak apabila $t < t_{(n_1 + n_2 - 2)} \cdot (1 - \alpha/2)$

Contoh :

Dua jenis pupuk buatan telah dipergunakan di tanah pertanian yang memiliki tingkat kesuburan, maupun keadaan iklim yang kurang lebih sama. Tujuan penggunaan pupuk di atas ialah untuk menguji apakah daya hasil salah satu jenis pupuk buatan tersebut betul-betul berbeda dari pada yang lain. Penyelidik memilih secara random 12 petak tanah pertanian dan memberinya dengan pupuk buatan pertama dan 12 petak tanah pertanian lainnya untuk diberi pupuk buatan kedua.

Hasil pertambahan padi dalam kilogram diberikan sbb. :

Hasil penggunaan yang pertama, x_{1i}	31	34	29	26	32	35	38	34	30	29	32	31
Hasil penggunaan yang kedua	26	24	28	29	32	26	31	29	31	29	32	28

Apakah hasil penggunaan yang pertama benar-benar berbeda dari yang kedua?

Penyelesaian : $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$ Data unpaired.

Andaikan $x_1 \sim N(\mu_1, \sigma^2)$ dan $x_2 \sim N(\mu_2, \sigma^2)$

Menurut hasil di atas maka H_0 harus ditolak apabila :

$$\left| \frac{\bar{x}_1 - \bar{x}_2}{\delta p \sqrt{\frac{1}{12} + \frac{1}{12}}} \right| > t_{(12+12-2)} ; 0,975$$

Disini $\bar{x}_1 = 31,75$, $\bar{x}_2 = 28,67$, $s_1^2 = 10,2045$, $s_2^2 = 6,0580$

Sehingga $\frac{\bar{x}_1 - \bar{x}_2}{\delta p \sqrt{\frac{1}{6}}} = 2,526$ sedangkan $t_{22, 0,975} = 2,074$

Jadi H_0 harus ditolak.

(iii) Soal :

Diketahui $x_1 \sim N(\mu_1, \sigma_1^2)$ dan $x_2 \sim N(\mu_2, \sigma_2^2)$

dimana σ_1^2 dan σ_2^2 tidak diketahui. Disajikan data bilang-

an-bilangan nyata $x_{11}, x_{12}, \dots, x_{1n}$ berupa random sample dari r.v. x_1 , serta bilangan-bilangan nyata $x_{21},$

x_{22}, \dots, x_{2n} berupa random sampel dari r.v. x_2 , dimana

x_{1i} dan x_{2i} merupakan ukuran-ukuran terhadap unsur ke 1

dalam dua macam keadaan/pada dua waktu, dlsb. (paired data).

$$H_0 : \mu_1 = \mu_2$$

Penyelesaian:

Didefinisikan $x = x_1 - x_2$, berarti $x \sim N(u, \sigma^2)$

dimana $u = u_1 - u_2$ dan $\sigma^2 = \sigma_1^2 + \sigma_2^2$

H_0 menjadi berupa : $u = 0$

Karena itu kita dapat memakai hasil (i) tersebut di atas.

Misalnya, bila H_1 ialah $u_1 \neq u_2$ maka H_0 harus ditolak

apabila $\left| \frac{\bar{x}}{\delta/\sqrt{n}} \right| > t_{(n-1)}, (1-\alpha/2)$

Contoh :

Kita sajikan data sebagai berikut,

x_{1i}	66	34	12	48	91	69	23	80	75	47
x_{2i}	54	26	11	59	70	42	25	58	83	67

dimana x_{1i} = nilai ujian akhir semester rata-rata mahasiswa i dalam kurikulum lama.

dan x_{2i} = nilai ujian akhir semester rata-rata mahasiswa i dalam kurikulum baru.

Uji hipotesa menghasilkan kesimpulan bahwa kurikulum baru merugikan para mahasiswa.

Pengujian hipotesa dua cara penggolongan tertentu yang tidak ada hubungannya/tidak saling mempengaruhi (independent).

Soal : Disajikan data bilangan-bilangan bulat positif.

$n_{11}, n_{12}, \dots, n_{1c}$ dimana n_{ij} = jumlah unsur
 $n_{21}, n_{22}, \dots, n_{2c}$ yang digolongkan ke dalam golongan
 $n_{R1}, n_{R2}, \dots, n_{Rc}$ i dalam cara penggolongan pertama
 dan digolongkan ke dalam golongan
 j dalam cara penggolongan kedua.

Cara II

	1	2	c	jumlah
1	n_{11}	n_{12}	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	n_{2c}	$n_{2.}$
.....
.....
.....
R	n_{R1}	n_{R2}	n_{Rc}	$n_{R.}$
jumlah	$n_{.1}$	$n_{.2}$	$n_{.c}$	n

Cara I

H_0 : cara-cara penggolongan itu independent.

Penyelesaian:

Bila H_0 benar maka jumlah unsur yang diharapkan terdapat di dalam kotak (i, j) ialah $J_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$

Selanjutnya diketahui bahwa, asal semua $J_{ij} \geq 5$

maka kira-kira $Q = \sum_{ij} \frac{(n_{ij} - J_{ij})^2}{J_{ij}} \sim \chi^2_{(R-1)(c-1)}$

dan Q cenderung lebih besar bila H_0 tidak benar.

\therefore Hitung Q dan tolak H_0 bila $Q > \chi^2_{(R-1)(c-1), (1-\alpha)}$

Catatan:

1. Asal R dan c tidak kecil maka pendekatan masih cukup teliti bila kurang dari pada 20% diantara $J_{ij} < 5$, asal tidak ada yang < 1 .

$$2. \text{ Bila } R = c = 2 \text{ maka } Q = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}}$$

3. Bila n kecil maka pendekatan tersebut kurang teliti. Ketelitiannya dapat dipertinggi kalau Q diganti dengan

$$Q' = \sum_{ij} \frac{(fn_{ij} - J_{ij} - \frac{1}{2})^2}{J_{ij}}$$

Contoh :

Seorang pengusaha akan menyelidiki apakah tingkat pendidikan mempengaruhi hasil pekerjaan bagi ke 200 pegawai yang dipekerjakan di pabriknya. Untuk tujuan tsb. dia mengadakan tiga golongan menurut tingkat pendidikan, demikian pula menurut hasil pekerjaan, untuk hasilnya lihat tabel yang berikut:

		Tingkat Pendidikan			
Hasil Pekerjaan		SD	SMP	SMA Ke atas	Jumlah
	Bagus	10	40	10	60
	Baik	30	30	20	80
	Sedang	10	30	20	60
	Jumlah	50	100	50	200

Dapatkah dikatakan bahwa pengaruh benar-besar ada?

Penyelesaian:

Tabel yang menunjukkan jumlah pegawai yang diharapkan di dalam setiap hasil penggolongan dua cara adalah sbb.:

15	30	15
20	40	20
15	30	15

Sehingga kita mendapatkan bahwa $Q = 17,5$

Sedangkan $\chi^2_{(3-1)(3-1);0,99} = \chi^2_{4;0,99} = 13,3$

Jadi H_0 ditolak apabila kita pakai level of significance 1%, artinya kita bisa yakin bahwa pengaruh tsb. benar - benar ada.

Pengujian hipotesa dua r.v. tertentu yang tidak ada hubungannya/ tidak saling mempengaruhi (korelasi) :

Soal: Diketahui bahwa r.v. x dan y bersama mempunyai distribusi bivariate normal, sehingga masing-masing sendiri mempunyai distribusi normal.

Disajikan data pasangan-pasangan bilangan nyata (x_1, y_1) (x_n, y_n) .

$H_0 : \rho = 0$, di mana ρ = koefisien korelasi dari x, y .
 $(-1 < \rho < 1, |\rho| = 1$ bila $y = ax$ di mana a = bilangan tetap).

Penyelesaian:

Bila H_0 benar maka $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{(n-2)}$

$$\text{dimana } r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Hitung t . Bila H_1 adalah " $\rho \neq 0$ " maka H_0 harus ditolak kalau $|t| > t_{(n-2)} ; (1-\alpha/2)$.

Bila H_1 adalah " $\rho > 0$ " maka H_0 harus ditolak kalau $t > t_{(n-2)} ; (1-\alpha/2)$.

Bila H_1 adalah " $\rho < 0$ " maka H_0 harus ditolak kalau $t < -t_{(n-2)} ; (1-\alpha/2)$.

Contoh:

Data yang pernah dikumpulkan mengenai tinggi badan anak perempuan dan anak laki-laki satu keluarga adalah sebagai berikut:

	No. urut keluarga										
	1	2	3	4	5	6	7	8	9	10	11
Laki-laki (x)	71	65	66	67	70	71	70	73	72	65	66
Perempuan (y)	69	64	65	63	65	62	65	64	76	59	62

Apakah dapat disimpulkan bahwa ada korelasi antara kedua variabel itu?

Penyelesaian:

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0 \quad (\text{Dikira-kira tak mungkin } \rho < 0)$$

Andaikan x dan y bersama mempunyai distribusi bivariat normal. Menurut hasil di atas maka H_0 harus ditolak apabila:

$$\frac{r \cdot \sqrt{9}}{\sqrt{1-r^2}} > t_{9;0,95}$$

Disini $r = \frac{\sum xy - \frac{1}{n} (\sum x) (\sum y)}{\sqrt{(\sum x^2 - \frac{1}{n} (\sum x)^2) (\sum y^2 - \frac{1}{n} (\sum y)^2)}} = \dots\dots = 0,558$

$$\text{Sehingga } \frac{3r}{\sqrt{1-r^2}} = 2,02$$

$$\text{Sedangkan } t_{9;0,95} = 1,833$$

Jadi H_0 ditolak, artinya korelasi dianggap ada.

Analisa Varian.

Analisa varian digunakan untuk pengujian hipotesa tentang pengaruh sebuah faktor terhadap suatu variabel. Faktor yang bersangkutan boleh "diskrit", misalnya hari, merek bahan yang dipakai orang, juga boleh "kontinu", misalnya temperatur, banyaknya air, lamanya memakai.

Soal:

Disajikan bilangan nyata dalam bentuk:

$$\begin{array}{l} x_{11}, x_{12}, \dots, x_{1n_1} \\ x_{21}, x_{22}, \dots, x_{2n_2} \\ \hline x_{a1}, x_{a2}, \dots, x_{an_a} \end{array} \quad \sum_i n_i = N$$

di mana data di baris pertama yang bersangkutan dengan tingkat 1 dari faktor,
data di baris kedua yang bersangkutan dengan tingkat 2 dari faktor, dst.

Diandaikan $x_1 \sim N(u_1, \sigma^2)$ bagi tingkat 1 dari faktor,
 $x_2 \sim N(u_2, \sigma^2)$ bagi tingkat 2 dari faktor, dst.

$H_0 : u_1 = u_2 = \dots = u_a$, artinya faktor tidak berpengaruh.

Penyelesaian:

Bila H_0 benar maka diketahui bahwa

$$F = \frac{\frac{1}{a-1} \sum_i n_i (\bar{x}_i - \bar{x})^2}{\frac{1}{N-a} \sum_{ij} (x_{ij} - \bar{x}_i)^2} \sim F_{(a-1), (N-a)} \quad \text{di mana}$$

$$\bar{x}_i = \frac{1}{n_i} \sum_j x_{ij} \quad \text{dan} \quad \bar{x} = \frac{1}{N} \sum_{ij} x_{ij}$$

dan F cenderung lebih besar bila H_0 tidak benar.

.. Hitung F dan tolak H_0 bila $F > F_{(a-1), (N-a); (1-\alpha)}$

Contoh: Pertambahan berat badan rata-rata sehari 3 kelompok babi yang masing-masing kelompok diberi tambahan vitamin dengan jumlah yang tertentu adalah sebagai berikut:

Tingkat vitamin (gm/kgm makanan)	Pertambahan berat (kg)		
0	1,52	1,56	1,54
10	1,63	1,57	1,54
20	1,44	1,52	1,63

Apakah tambahan vitamin itu berpengaruh?

Penyelesaian:

Tingkat faktor	x_{ij}			$\Sigma x_{i.}$	$\bar{x}_{i.}$
1	1,52	1,56	1,54	4,62	1,54
2	1,63	1,57	1,56	4,74	1,58
3	1,44	1,52	1,63	4,59	1,53

$$\Sigma x_{..} = 13,95, \quad \bar{x} = 1,55$$

$$\Sigma_i n_i (\bar{x}_i - \bar{x})^2 = \Sigma_i n_i \bar{x}_i^2 - N \bar{x}^2 = \dots = 0,0042$$

$$\begin{aligned} \Sigma_{i,j} (x_{ij} - \bar{x}_i)^2 &= \Sigma_{i,j} (x_{ij} - \bar{x})^2 - \Sigma_i n_i (\bar{x}_i - \bar{x})^2 \\ &= \Sigma_{i,j} x_{ij}^2 - N \bar{x}^2 - 0,0042 = 0,0232 \end{aligned}$$

$$F = \frac{\frac{1}{3-1} \Sigma_i n_i (\bar{x}_i - \bar{x})^2}{\frac{1}{9-3} \Sigma_{i,j} (x_{ij} - \bar{x}_i)^2} = 0,538$$

Menurut hasil di atas maka H_0 harus ditolak apabila :

$$F > F_{(3-1);(9-3); 0,95} \quad \text{Tetapi} \quad F_{2;6;0,95} = 5,14$$

Jadi H_0 diterima, artinya tambahan vitamin itu belum pasti berpengaruh.

LATIHAN STATISTIK

1. Sebuah sample yang terdiri dari 15 kaleng cat memiliki isi berat kotor seperti yang diberikan di bawah ini. Isi berat dinyatakan dalam kg per kaleng.

1,21	1,21	1,23	1,20	1,21
1,24	1,22	1,24	1,21	1,19
1,19	1,18	1,19	1,23	1,18

Dika kita digunakan level of significance sebesar 1%, yakinkah bahwa populasi cat dalam kaleng secara rata-rata memiliki isi berat kotor 1,2 kg per kaleng? Berilah evaluasi Saudara.

2. Dua buah sample yang masing-masing dipilih dari 2 populasi menghasilkan data seperti berikut:

Sample	n	d.f.	\bar{x}	$\sum (x_i - \bar{x})^2$
1	15	14	21	1.224
2	9	8	29	756
		22		1.980

Apakah ada alasan guna mempercayai pendapat bahwa rata-rata populasi pertama lebih kecil dari rata-rata populasi kedua? Penggunaan level of significance $\alpha = 0,05$

3. Daftar di bawah ini menyajikan hasil observasi penduduk kota yang diklasifikasi atas dasar atribut warna mata dan warna rambut.

Warna mata \ Warna rambut	Warna rambut		
	Hitam	Coklat	
Hitam	32	12	44
Coklat	14	22	36
Lain-lain	6	9	15
	52	43	95

Apakah ada alasan guna menganggap bahwa atribut penduduk kota di atas independen?

Pergunakan $\alpha = 0,10$ dan beri evaluasi Saudara.

4. Suatu koefisien korelasi dari 0,50 yang diperoleh dalam suatu sample dari 27 pasang. Dapatkan koefisien korelasi itu dipandang mempunyai perbedaan yang berarti dari 0, berdasarkan atas 5% level of significance?
5. Dapatkan ke 5 sample di bawah ini dipandang berasal dari populasi dengan macam yang sama berdasarkan 5% level of significance?

A	B	C	D	E
10	12	9	10	10
4	10	4	6	6
6	13	4	9	4

4. PENGUMPULAN DATA

Setiap penelitian memerlukan data. Data ini sangat diperlukan baik untuk mengembangkan hipotesa maupun untuk menguji hipotesa.

Seorang peneliti yang tajam persepsinya terhadap data yang dibacanya dari berbagai sumber pustaka seperti buku-buku teks, buletin, abstrak, buku sumber data (*data source*), maupun berkala ilmiah seperti *periodicals* atau *journals* dapat mengembangkan hipotesa tentang masalah yang sedang ditelitinya. Sedangkan untuk menguji hipotesanya peneliti harus mengumpulkan data melalui berbagai cara, antara lain: studi lapangan, studi laboratorium, penyebaran kuesioner, wawancara, atau melakukan eksperimentasi.

Data merupakan komponen penting dalam penelitian, sebab dari himpunan data dapat disimpulkan informasi-informasi yang diperlukan. Oleh karena itu seorang peneliti perlu mengenal makna dan fungsi data.

Data dan karakteristiknya ditentukan oleh berbagai faktor. Faktor pertama adalah sasaran pengamatan. Oleh karena menurut sasarannya terdapat dua macam pengamatan, yaitu pengamatan kuantitatif dan pengamatan atributif atau kualitatif maka dalam penelitian kita mengenal dua macam data, data kuantitatif dan data kualitatif. Variasi di antara dua macam tadi disebut data semi kuantitatif. Jika seseorang peneliti mengukur karakteristik sampel, misalnya tinggi, berat, atau volume suatu obyek maka ia akan memperoleh data kuantitatif. Dan terhadap data tersebut dikenakan satuan seperti cm, kg, atau liter. Jika pengamatan peneliti ditujukan untuk mengungkapkan atribut suatu obyek ia akan memperoleh data kualitatif yaitu informasi mengenai atribut obyek yang diselidiki, misalnya laki-laki atau perempuan, kawin atau belum kawin.

Suatu himpunan data yang berasal dari sistem kualitatif dapat pula dinyatakan dengan sistem kuantitatif dengan menyajikan angka-angka urutan (*rank-order*) tentang tingkatan kualitas atributnya, data semacam ini merupakan data semi kuantitatif.

Faktor kedua adalah kontinuitasnya karakteristik sampel yang diamati.

Misalnya mengenai usia seseorang, berat dan tinggi badan bila diamati akan menghasilkan data kuantitatif yang kontinu yang analisisnya menuntut cara tersendiri. Sebaliknya yang menyangkut frekuensi suatu kejadian, misalnya berapa jumlah mahasiswa yang gagal studi, berapa penduduk yang menderita busung-lapar, jika diamati akan menghasilkan data yang diskrit.

Sehubungan dengan macam data dan karakteristiknya tersebut di atas seorang peneliti dituntut untuk mempertimbangkan macam data, cara pengumpulan, dan disain analisisnya lebih dahulu sebelum melangkah ke tahap pengumpulan data.

Berdasarkan orang atau badan yang mengumpulkannya, data penelitian ada yang disebut data interen dan ada data yang disebut data eksteren. Data interen adalah data yang diambil atau dikumpulkan oleh peneliti atau badan riset yang melakukan aktivitas penelitian itu sendiri. Bila peneliti atau badan riset tersebut mengambil datanya dari data yang telah dikumpulkan oleh peneliti atau badan peneliti lain maka data tersebut bagi peneliti dan badan peneliti pemakai data disebut data eksteren.

Misalnya suatu Tim Peneliti ingin meneliti sikap atau reaksi guru terhadap Kurikulum SMA 1975. Jika tim tersebut dalam mengumpulkan datanya secara langsung terjun ke sekolah-sekolah, mengamati sendiri, melakukan wawancara, untuk memperoleh data tentang sikap dan reaksi guru terhadap kurikulum tersebut maka data yang dikumpulkan dapat dikategorikan sebagai data interen. Tetapi jika tim tersebut tidak melakukan pengumpulan data sendiri secara langsung dan hanya mengumpulkan informasi-informasi dari sumber lain, seperti berita surat kabar, pernyataan pers, pejabat departemen, majalah, laporan-laporan dinas, maka data yang dikumpulkannya dapat dikategorikan sebagai data eksteren.

Dalam penelitian data interen merupakan data pokok, sedangkan data eksteren dianggap sebagai data tambahan yang lebih banyak berfungsi sebagai data pembanding.

Mengenai jenis data ada klasifikasi yang membedakan data primer dan data sekunder.

Dalam buku *Modern Business Statistics* karangan John E. Freud dan Frank J. Williams terbitan Prentice-Hall, New Jersey halaman 468 tercantum data:

The 1958 Federal Budget :

Expenditures: 60,3 %	Security programs
10,3 %	Interest
7,0 %	Veterans
7,0 %	Agriculture
15,4 %	All other

(Source : Associated Press)

Bagi Associated Press yang mengumpulkan data itu untuk kepentingan mengungkapkan permasalahan adalah data primer. Akan tetapi bagi orang lain, misalnya anggota parlemen, atau badan peneliti lain yang mengutip data tersebut untuk kepentingan penelitiannya adalah data sekunder.

Departemen, Universitas, Badan Riset atau badan-badan swasta tertentu dapat menerbitkan data hasil penelitiannya secara berkala. Di Indonesia ada suatu badan pemerintah yang bertugas untuk mengumpulkan, mengolah dan menerbitkan data, yaitu Biro Pusat Statistik di Jakarta. Peneliti yang berminat dapat memperoleh data sekunder untuk menunjang data primernya. Beberapa data yang diterbitkan oleh biro itu, antara lain:

- (1) Luas Tanaman, Produksi dan Persediaan Tanaman-tanaman Terpenting;
- (2) Ekspor Menurut Jenis Barang;
- (3) Impor Menurut Jenis Barang;
- (4) Statistik Konjunktur;
- (5) Perdagangan Antar Pulau;
- (6) *Statistical Pocket Book of Indonesia*;

Badan lain yang hampir setiap tahun menerbitkan data semacam itu di Indonesia adalah Bank Indonesia yang terbitannya berjudul "Laporan Bank Indonesia".

Pengumpulan data merupakan salah satu langkah penelitian yang perlu dipikirkan dengan cermat. Beberapa pertimbangan perlu dilakukan lebih dahulu, antara lain:

- (a) Macam data apakah yang perlu dikumpulkan;
- (b) Bagaimanakah cara yang efisien dan efektif untuk mengumpulkannya;
- (c) Manakah sumber data yang sesuai untuk memperoleh data yang diperlukan.

Mengenai cara pengumpulan data sangat dipengaruhi oleh maksud, sifat, dan tujuan penelitian. Data dapat dikumpulkan dengan satu cara pengumpulan data, misalnya dengan wawancara, dengan menyebarkan daftar pertanyaan atau kuesioner, dengan pengukuran-pengukuran, dengan investigasi laboratorium, atau dengan eksperimentasi. Akan tetapi seorang peneliti dapat juga memakai kombinasi berbagai cara pengumpulan data.

Besarnya biaya penelitian, efisiensi, dan efektivitas penelitian, waktu dan fasilitas yang tersedia serta kemampuan peneliti untuk mengumpulkan data, dan kelangkaan sumber data sangat mempengaruhi cara pengumpulan data.

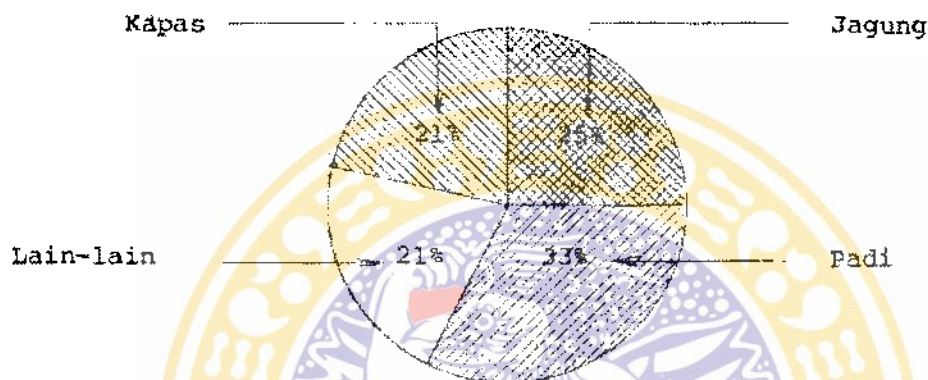
Tampaknya pengumpulan data dengan mengirimkan daftar pernyataan sangat murah biayanya. Tetapi dengan cara itu kadang-kadang kurang efektif, karena sangat besar kemungkinannya hanya sedikit responden yang mengembalikan daftar pertanyaan terjawab kepada peneliti, sehingga data yang diperoleh menjadi kurang meyakinkan. Jika diperhitungkan dengan anggaran yang dikeluarkan, misalnya ongkos cetak, upah pembantu teknis, biaya administrasi, dan porto dengan sedikitnya respon maka biaya per unit data menjadi mahal.

Oleh karena faktor-faktor tersebut maka sebelum melangkah ke pengumpulan data peneliti sebaiknya mendasarkan teknik pengumpulan datanya pada suatu rancangan atau disain yang baik dan fisibel. Disain yang telah terancang baik dapat dipakai sebagai pegangan selama terjun ke lapangan sehingga dapat dihindari adanya keborosan dan tidak meyakinkannya data.

Di samping disain pengumpulan data seorang peneliti juga harus memikirkan bagaimana cara penyajian datanya jika data itu telah terkumpul. Seringkali terjadi bahwa hasil penelitian mengalami kegagalan, sia-sia atau tidak mencapai tujuan hanya karena teknik penyajian datanya yang tidak efisien.

Tidaklah terlalu berlebihan jika dikatakan bahwa metode penyajian fakta lebih mudah dipahami dari pada fakta itu sendiri. Barangkali yang paling cepat meyakinkan dan yang paling mudah dipahami adalah penyajian data dalam bentuk diagram atau gambar lambang, seperti contoh berikut.

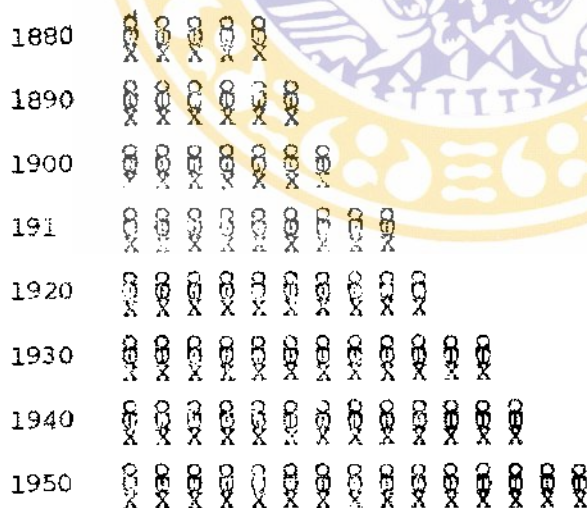
(1) Penyajian data dalam bentuk diagram.



Produksi yang terbesar ialah

(2) Penyajian data dalam bentuk gambar lambang.

Penduduk Amerika Serikat 1880 - 1950.



EACH SYMBOLE = 10 MILLION PEOPLE

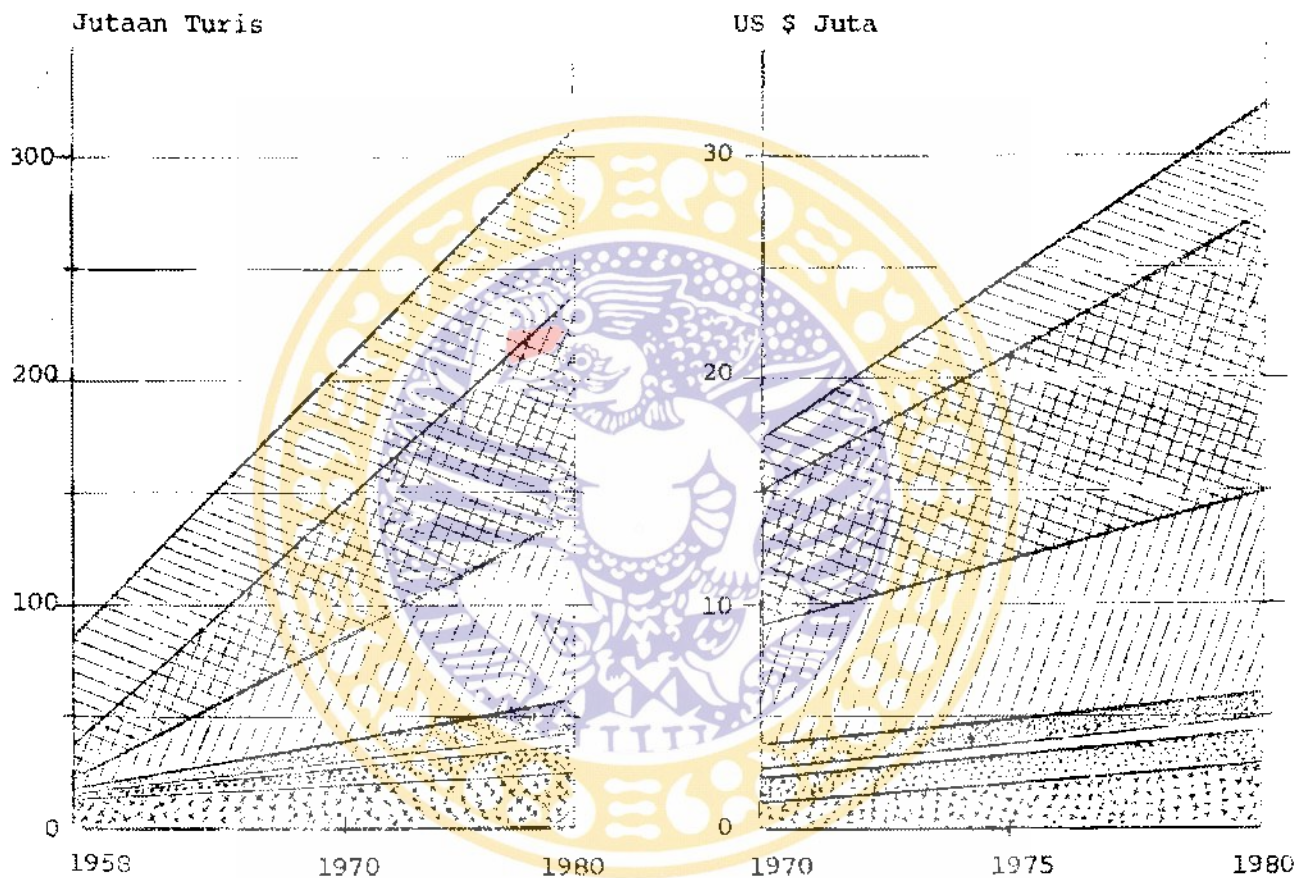
(Source : 1956 Statistical Abstract of the United States).

Ini berarti:

- 2.1. Penduduk A.S. tahun 1900 ialah orang.
- 2.2. Penduduk A.S. tahun 1950 ialah orang.

Disamping visualisasi data seperti tersebut di atas data juga dapat disajikan dalam bentuk kurva atau gambar grafik, misalnya:

(3) Kurva Pariwisata & Pengeluaran Wisatawan Internasional.



Keterangan :








-  Amerika Utara
-  Eropa Barat
-  Wilayah Laut Tengah
-  Eropa Timur
-  Afrika
-  Asia dan Oceania
-  Amerika Selatan, Tengah dan Karibia

TABLE RISING ENROLLMENT RATIOS IN SAMPLE OF DEVELOPING COUNTRIES.

COUNTRIES	PRIMARY LEVEL (percentage of 5 - 14 years olds)				SECONDARY LEVEL (percentage of 15 - 19 years olds)			
	1950	1955	1960	1963	1950	1955	1960	1963
AFRIKA								
Cameroon	25	37	57	79	0,7	2	5	8
Niger	1	2	4	7	0,1	0,2	0,5	0,8
Nigeria	16	27	35	31	1	2	5	6
Tunisia	18	26	44	57	9	11	15	20
ASIA								
Afganistan *)	3	4	5	9	0,5	0,7	1	3
Indonesia **)	29	39	42	45	3	7	9	10
Korea, Re- public of	53	54	60	69	20	36	32	38
Pakistan	16	19	22	25	15	15	16	22
LATIN AMERIKA								
Argentina	66	71	69	68	21	28	32	37
Bolivia	24	34	38	49	9	12	15	19
Brazil	28	33	45	50	10	12	18	25
Venezuela	40	44	70	72	6	11	27	34

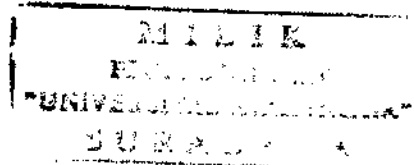
Source: UNESCO, Statistical Yearbook, 1965 (Paris, 1966), pp. 177-137.

*) Public education only.

**) Not including West Irian.

5. TEKNIK SAMPLING

5.1. Pendahuluan



Bila seorang peneliti ingin menyelidiki masalah pengangguran di Jawa Timur, pertama-tama yang perlu difikirkan adalah karakteristik dan batasan variabel dari populasi yang akan diselidiki, dalam hal ini adalah pengangguran.

Siapa di Jawa Timur ini yang termasuk pengangguran? Bila seseorang hanya dapat bekerja 1 jam per hari, apakah dia dapat digolongkan pengangguran? Seorang petani yang hanya mempunyai sebidang sawah; tiga bulan setelah sawahnya ditanami, dia tidak bekerja, kecuali menunggu hasilnya. Apakah pak tani yang disebut tadi adalah juga seorang pengangguran?

Jelaslah bahwa pada setiap penelitian batasan yang jelas terhadap variabel dan populasi yang akan diteliti perlu ditegaskan sebelum melangkah ke pengumpulan data.

Apakah populasi itu? Populasi ialah semua obyek yang akan diselidiki. Kalau kita menyelidiki masalah pengangguran di Jawa Timur, maka populasinya adalah semua pengangguran di Jawa Timur.

Populasi dapat dibedakan atas :

- (1) Populasi yang terhingga yaitu bila jumlah anggota populasi terbatas dan dapat dihitung, misalnya :
 - a. Jumlah mahasiswa IKIP Surabaya tahun 1975
 - b. Pegawai Negeri di Jawa Timur tahun 1975
 - c. Jumlah penduduk Kotamadya Surabaya tahun 1976.
- (2) Populasi yang tak terhingga yaitu populasi yang jumlah anggotanya tak dapat dihitung, misalnya:
 - a. Volume air Laut Jawa
 - b. Batang padi PB 5 yang tumbuh di Jawa Timur tahun 1975?
 - c. Populasi ikan paus yang masih hidup di Samudera Indonesia?

Di dalam praktek para peneliti sering mempunyai maksud menyelidiki karakteristik populasi yang tak berhingga, sedangkan biaya penelitian umumnya sangat terbatas. Oleh karena itu maka peneliti cenderung untuk hanya mengamati karakteristik sampel tertentu dan menggunakan data yang diperolehnya untuk mengestimasi karakteristik populasinya. Untuk mencapai tujuan itu perlu disusun disain penelitian yang tepat.

Di dalam disain itulah kita dapat membaca, seberapa jauh efisiensi penelitian itu nantinya. Sudahkah sepadan antara hasil penelitian tersebut dengan biaya yang dikeluarkan? Kalau biaya sangat terbatas, strategi apakah yang perlu digariskan oleh si peneliti agar dapat juga diperoleh hasil yang masih dapat dipercaya (reliable)? Dalam keadaan biaya sangat terbatas strategi yang biasanya diambil ialah *strategi sampling*. Artinya si peneliti tidak mengamati seluruh anggota populasi, melainkan cukup dengan sebagian kecil saja. Bagian dari populasi yang anggotanya dapat dikuasai, disebut sampel. Mengira-ira karakteristik populasi hanya dengan berdasarkan analisa data karakteristik sampel disebut metode *sampling*.

Ukuran karakteristik populasi disebut parameter, misalnya harga rata-rata hitung, standard deviasi, atau mode dari populasi. Parameter ini dapat diperoleh dengan menghitung atau mengukur karakteristik yang terdapat pada anggota populasi satu per satu. Cara menghitung anggota seperti ini disebut *sensus*. Tetapi di dalam praktek parameter ini lebih sering di kira-kira dengan cara *sampling*.

Mengapa di dalam praktek kita lebih sering menghitung parameter dengan jalan *sampling*? Sudah bisakah dipercaya bahwa hasil hitungan statistik dari sampel sudah mewakili karakteristik populasi yang akan diteliti?

Pertanyaan ini dapat dijawab "ya", bila *sampling*nya telah memenuhi syarat yaitu :

- a. Representatifnya sampel, artinya anggota sampel yang diteliti memiliki karakteristik yang serasi dengan karakteristik populasi yang akan diteliti.
- b. Adekuat atau cukupnya jumlah sampel.

Di bawah ini diberikan satu contoh teknik sampling yang tidak memenuhi syarat sampling yang baik. Tahun 1936, majalah "*Literary Digest*" di Amerika membuat ramalan bahwa pemilihan umum kepresidenan tahun itu akan dimenangkan oleh Partai Republik. Ramalan itu didasarkan atas hasil wawancara (*polling*) sejumlah sampel yang sangat besar. Anggota sampel yang diwawancarai diambil dari daftar buku telepon dan dari buku registrasi pemilik mobil. Hasil *polling* diumumkan secara luas oleh majalah tersebut. Bagaimana hasilnya setelah pemilihan umum usai? Ternyata kursi kepresidenan dimenangkan oleh Roosevelt (Demokrat). Ramalan majalah "*Literary Digest*" meleset sama sekali. Yang diramalkan adalah Partai Republik, kenyataannya dimenangkan oleh Partai Demokrat. Di mana letak kesalahan strategi sampling dari the "*Literary Digest*" tersebut? Bukankah besar sampel telah mencukupi? Besar sampel telah mencukupi, hanya sampelnya tidak representatif, sebab pada waktu itu hanya anggota masyarakat Amerika yang berada saja yang rumahnya memiliki telepon dan memiliki mobil. Jadi sampelnya hanya mewakili lapisan masyarakat yang berada saja. Sedang pendapat golongan masyarakat bawah tidak diketahui, sebab tidak ikut diwawancarai, justru Roosevelt pada waktu itu mendapat dukungan luas dari lapisan masyarakat pekerja yaitu lapisan masyarakat yang tak berada?

5.2. Representatif - tidaknya sampel.

Kesalahan teknik sampling seperti yang dibuat oleh the "*Literary Digest*" itu adalah satu kesalahan sampling yang klasik, kesalahan yang ternyata membawa bencana bagi majalah itu sendiri, kesalahan yang terletak pada pengambilan sampel yang tidak representatif.

Kalau sampel yang diambil sudah representatif, berarti teknik sampling akan :

- a. Menyederhanakan analisa data
- b. Menghemat biaya, tenaga, dan waktu
- c. Dapat menghitung hal-hal yang tidak dapat diukur atau dihitung.

Ketiga kebaikan teknik sampling itulah yang menyebabkan mengapa di dalam praktek kita hampir selalu bekerja dengan sampel. Teknik sampling inilah yang merupakan bagian penting setiap penelitian.

Teknik sampling manakah yang dapat memenuhi persyaratan? dan bagaimanakah rancangan atau disain sampel yang perlu dikembangkan?

Disain sampel adalah suatu rancangan yang jelas, yang disusun lengkap sebelum pengumpulan data dimulai, yang bertujuan untuk mengambil sampel dari suatu populasi yang karakteristiknya akan diselidiki.

Sampling yang baik adalah sampling yang anggota sampelnya diambil secara random. Anggota sampling yang diambil secara random (*random sampling*) artinya bila setiap anggota sampel mendapat kesempatan yang sama untuk dipilih menjadi anggota sampel.

Dalam penelitian dikenal beberapa macam disain atau rancangan sampel (*sampling schemes*), antara lain :

- (1) Sampling bertingkat (*stratified sampling*);
- (2) *Cluster sampling*;
- (3) Sampling sistematis;

yang semuanya dapat dikategorikan ke dalam *random sampling*.

5.2.1. *Sampling bertingkat.*

Adakalanya populasi itu bersifat homogen dan adakalanya heterogen. Contoh populasi yang homogen ialah semua padi PB 5 di Jawa Timur. Contoh populasi yang heterogen ialah pemeluk agama di Indonesia.

Populasi yang heterogen dapat dibagi-bagi atas sub-populasi yang disebut strata (tunggal : *stratum*), misalnya mengenai populasi pemeluk agama di Indonesia dapat dibagi menjadi strata pemeluk agama Islam, Kristen, Hindu. Populasi jenis padi dapat dibagi menjadi strata jenis padi Bengawan, PB 5, PB 8, Synthia, Sri Gadis, Dewi Ratih dan sebagainya.

Dalam hal ini populasi yang heterogen dibagi menjadi sub populasi yang homogen.

Kalau populasinya heterogen, maka sampelnya diambil dari masing-masing stratumnya. Karena itu teknik sampling semacam itu disebut sampling bertingkat (*stratified sampling*).

Kalau besar sampel yang diambil sebanding dengan besarnya sub populasi, maka teknik sampling yang demikian disebut sampling bertingkat sebanding (*proportional stratified sampling*).

Adakalanya populasi dibagi-bagi atas strata geografis. Dan masing-masing stratum dibagi atas karakteristik penduduk dalam satu stratum. Teknik sampling semacam ini disebut sampling stratifikasi silang (*cross-stratification sampling*). Misalnya kita ingin mengetahui sikap bangsa Indonesia terhadap program Keluarga Berencana. Untuk ini wilayah Indonesia dibagi-bagi menjadi beberapa daerah geografis, misalnya Bali, Jawa Timur, Jawa Tengah dan lain-lainnya. Kemudian penduduk masing-masing populasi dibedakan lagi atas masing-masing penduduk beragama Islam, Kristen, dan Hindu. Di sini tampak penduduk yang beragama Islam dan Kristen hampir ada pada setiap sub populasi (sub populasi menyilang). Karena itu teknik sampling semacam ini disebut sampling stratifikasi silang.

5.2.2. *Cluster sampling.*

Pada *cluster sampling* (sampling berkelompok), populasinya dibagi-bagi menjadi sub populasi yang tidak selalu homogen. Dalam hal ini beberapa kelompok (*clusters*) saja yang menjadi anggota sampel. Kelompok (*clusters*) yang menjadi anggota sampel ini diambil secara random. Bila kelompok (*clusters*) itu adalah daerah geografis, maka teknik sampling ini disebut *area sampling*.

5.2.3. *Sampling sistematis.*

Pada sampling sistematis anggota populasi diberi nomor urut. Kalau jumlah anggotanya 1.000 dan sampelnya 10%, dan anggota no. 4 dijadikan anggota sampel pertama maka anggota sampel lainnya terdiri dari anggota bernomor 14, 24, 34, 44, 54, 64, 74, dan seterusnya sehingga diperoleh 100 sampel.

4.3. Mekanisme Pengambilan Secara Random.

Karena dalam praktek, kita hampir selalu bekerja dengan *random sample*, maka perlu dibicarakan teknik pengambilan sampel secara *random*.

Misalkan di Kotamadya Surabaya ada 516 restoran, baik yang besar maupun yang kecil sederhana. Kalau kita ingin mengetahui sifat ke 516 restoran itu (misalnya rata-rata pengunjung/restoran/hari) kita tidak perlu menugaskan satu orang petugas pada setiap restoran untuk mencatat pengunjung per harinya.

Karena kalau demikian, biaya akan menjadi boros, dan pekerjaannya kurang efisien.

Biaya untuk membayar 516 petugas untuk mengumpulkan data sangat besar. Di sinilah tampak betapa pentingnya kita bekerja dengan sampel, khususnya dengan pengambilan sampel secara *random*.

Ada beberapa cara mengambil sampel secara *random*, antara lain :

5.3.1. *Dengan cara lot (undian).*

Caranya sebagai berikut : Buatlah daftar nama, alamat, dan nomor restoran secara urut. Buatlah daftar ini di dalam daftar populasi. Kemudian buatlah 516 gulungan lot (yang masing-masing berisi satu nomor restoran. Tentukan besar sampelnya, misalnya 10.

Sekarang ambilah 10 gulungan lot secara *random*. Setelah dibuka ternyata sampel pada kertas lot bernomor restoran 17, 18, 41, 50, 433, 496, 497, 503, 509, 608. Berarti kita harus

menyelidiki hanya 10 restoran yang nomornya seperti ini. Kita harus konsekwen mengamati 10 restoran itu, walaupun letaknya terpencar berjauhan. Janganlah mengelak dan akhirnya hanya mengambil restoran yang dekat-dekat saja.

5.3.2. *Radom sistematis.*

Caranya sebagai berikut : Kita sudah mempunyai daftar nama, alamat, dan nomor urut restoran. Dan besarnya sampel sudah ditetapkan, yaitu 10. Besar sampel $\frac{10}{516} \times 100\% = 2\%$ (atau $\frac{1}{50}$). Buatlah gulungan kertas lot yang bernomor urut 1 s/d 50. Ambil secara random hanya satu gulungan lot saja. Misalnya diperoleh no. 4, maka 10 restoran yang akan menjadi anggota sampel ialah restoran yang bernomor urut : 4, 54, 104, 154, 204, 254, 304, 354, 404, 454.

5.3.3. *Dengan tabel bilangan random.*

Pada halaman 76 & 77 disajikan contoh "Tabel Bilangan Random. Tabel ini memuat Heretan bilangan asli (0 - 9) dalam kolom-kolom. Di dalam tabel tercantum 40 kolom bilangan vertikal dan 2 x 25 kolom bilangan horisontal.

Kalau kita bekerja dengan sebuah sampel yang besarnya 10 dari satu populasi restoran yang besarnya 516, seperti contoh di atas, dan kita ingin mengambil sampel random itu dengan pertolongan "Tabel Bilangan Random" maka caranya adalah sebagai berikut : (Periksa tabel, perhatikan nomor-nomor bergaris bawah).

a. Ambil 3 kolom bilangan vertikal secara random (sebab 516 terdiri dari 3 angka).

Misalnya kita mengambil kolom bilangan vertikal nomor 9, 10 dan 11 dan dimulai dengan kolom horisontal 11.

Maka di dalam tabel itu kita dengan mudah menemukan bilangan : 480, 029, 286, 505, 177, 375, 001, 073, 430, dan 016.

- b. Nomor atau bilangan yang lebih tinggi dari 516 ditinggalkan misalnya no. 667, 801 dan sebagainya, sebab di Kotamadya Surabaya hanya ada 516 restoran. Jadi pasti tidak ada restoran yang nomor urutnya 667 atau 801.
- c. Kalau 2 nomor lama atau lebih yang semuanya lebih kecil dari 516, maka hanya satu nomor sajalah yang dipakai.

Pertanyaan:

Kalau kita memilih kolom no. 1, 2, 3, dan dimulai dari kolom horizontal no. 1, maka 10 restoran yang manakah yang seharusnya dijadikan anggota sampel?

Keuntungan dan Kerugian Mekanisme ini.

Cobalah bayangkan keuntungan-kerugian dan efisiensi kerja 3 mekanisme di atas bila :

- a. Populasinya tidak 516 tetapi 5783.
Mekanisme yang mana yang paling praktis?
- b. Populasinya hanya 37.
Mekanisme yang mana yang paling praktis?

Pengalaman kerja saudara sangat menentukan pilihan alternatif di atas. Di samping itu setiap pilihan sifatnya subyektif. Karena itu di sini tidak menjawab pilihan mana yang paling efisien dan paling praktis.

4.4. Keseeksamaan sampel.

Keseeksamaan sampel mencukupi atau menjawab pertanyaan cukup-tidaknya besar sampel untuk mewakili. Apakah sampel masih terlalu kecil? Ataukah sudah layak dan cukup?

Andaikan kita telah memiliki data sensus tentang rata - rata pengunjung restoran dari 516 restoran di Kotamadya Surabaya itu adalah 68 orang per restoran per hari. Marilah kita bayangkan bersama bagaimana kesimpulan kita bila :

- a. Kita mengambil sampel 5 restoran besar. Apakah rata-rata pengunjung pada sampel ini akan juga 68 orang per hari?
- b. Kalau kita mengambil sampel 5 restoran sekali lagi, sekali lagi dan sekali lagi, apakah rata-ratanya tetap sama 68 orang?
- c. Kalau kita mengambil sampel dengan jumlah restoran 10 atau 25 atau 100 atau 300 apakah rata-rata pengunjungnya pada sampel-sampel tersebut juga tetap 68 orang per restoran per harinya?

Contoh-contoh di atas bertujuan menyederhanakan persoalan. Di dalam praktek kita tidak selalu mengetahui besarnya populasi.

Di samping itu kita hampir selalu hanya mengambil satu sampel dari populasi yang tidak kita ketahui. Dalam hal seperti ini kita tidak selalu harus menggariskan strategi dan teknik sampling yang paling seksama. Kita harus mempunyai ukuran kemungkinan salahnya tafsiran sampling ini. Ahli-ahli statistik telah merumuskan besarnya kesalahan statistik, dengan berbagai ukuran, khususnya dalam kesalahan estimasi sampling ini disebut kesalahan standard (*standard-error*).

Cara pengukurannya adalah sbb.: *)

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$\sigma_{\bar{x}}$ = kesalahan standard dari estimasi rata-rata sampel terhadap populasi.

σ = standard deviasi populasi.

n = besar sampel.

*) Untuk perhitungan ini dan selanjutnya disarankan untuk membaca sekali lagi statistik deskriptif, statistik inferensial, dan menghitung kemungkinan (*probability*) atas dasar kurva normal.

Di dalam praktek kita tidak selalu mengetahui besarnya standard deviasi dari populasi (σ). Karena itu yang dipakai adalah standard deviasi sampel (s).

Dari rumus kesalahan standard ini kita dapat menarik kesimpulan bahwa makin besar sampel, makin kecil kesalahan standard estimasi.

Contohnya : Rata-rata pengunjung restoran di Kotamadya Surabaya adalah 68 orang per restoran per harinya dan standard deviasinya 3,44 orang. Ada 3 orang peneliti sosial (*social research workers*) masing-masing mengambil sebuah sampel random yang besarnya 5, 10, dan 25. Berapa besar kesalahan standard estimasi masing-masing peneliti tadi?

Jawab : Kesalahan standard ($\sigma_{\bar{x}}$) si peneliti A = $\frac{3,44}{\sqrt{5}} = 1,54$
 Kesalahan standard ($\sigma_{\bar{x}}$) si peneliti B = $\frac{3,44}{\sqrt{10}} = 1,09$
 Kesalahan standard ($\sigma_{\bar{x}}$) si peneliti C = $\frac{3,44}{\sqrt{25}} = 0,69$

Di dalam praktek kita hampir selalu bertanya seberapa besar sampel yang kita ambil? Tidak ada jawaban pasti untuk pertanyaan ini, apalagi besar populasi yang sedang kita hadapi tidak kita kenal. Dan populasi yang kita tidak kenal itulah yang justru kita selidiki.

Bila populasi yang kita selidiki itu kita ketahui, besar sampel yang dipakai dapat ditentukan dengan menggunakan rumus Parten :

$$N_s = \left(\frac{\sigma Z}{T} \right)^2$$

N_s adalah besar sampel yang layak dan cukup, bila parameter populasi telah diketahui. Karena perhitungannya adalah termasuk statistik inferensial, maka di sini tidak dibicarakan lebih lanjut. Dan lagi kita hampir selau menyelidiki populasi yang parameternya tidak kita ketahui sebelumnya.

SAMPLE PAGE OF RANDOM NUMBERS *

Twenty-seventh Thousand

	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40
1	29 69	05 96	07 99	56 72	57 45	44 28	70 30	00 69	94 05	59 93
2	44 14	89 31	38 89	04 66	38 42	79 00	57 41	01 48	86 12	97 08
3	40 64	84 72	35 39	88 73	92 01	35 62	56 26	31 50	62 61	53 60
4	46 96	38 41	53 66	41 32	51 91	97 94	58 28	11 46	58 80	20 79
5	93 59	07 05	58 89	90 94	08 97	89 17	04 08	02 37	37 88	83 84
6	14 55	32 12	51 49	28 37	95 62	64 61	53 81	05 83	75 50	45 42
7	84 92	59 57	52 52	71 08	94 38	31 61	99 19	48 09	53 43	10 17
8	57 48	69 92	22 71	97 59	57 53	07 05	79 68	98 13	13 92	47 73
9	92 43	03 74	24 54	63 01	45 01	84 14	88 83	14 70	23 80	88 61
10	66 26	06 22	46 50	51 37	66 79	13 46	82 29	36 15	28 94	51 80
11	34 33	36 05	48 04	43 28	94 89	42 00	50 81	45 12	25 97	00 00
12	11 16	92 13	66 71	23 57	66 65	58 45	79 97	54 04	58 24	76 22
13	24 98	49 31	80 11	70 43	83 64	51 29	98 11	84 00	73 31	34 82
14	00 63	87 95	02 90	50 58	76 21	62 77	17 16	48 47	34 81	64 19
15	33 05	54 37	28 69	55 55	83 30	97 82	48 19	77 58	09 84	66 85
16	60 61	08 49	50 52	91 94	86 46	28 05	69 83	35 88	30 33	92 31
17	66 51	16 28	79 45	01 45	99 28	31 36	42 84	67 55	74 13	23 47
18	39 72	97 77	68 79	95 80	90 13	96 75	96 46	73 30	78 52	13 34
19	30 85	20 65	17 78	65 50	64 81	17 75	17 98	74 43	28 55	97 65
20	85 45	39 53	37 59	26 97	54 78	64 75	67 78	92 87	32 82	55 53
21	55 86	15 01	37 96	10 81	73 50	10 95	07 70	00 36	80 78	85 96
22	79 90	44 95	65 31	60 23	87 45	20 49	06 50	27 89	76 77	17 55
23	37 71	95 08	00 15	33 41	88 52	37 46	29 53	78 97	43 03	80 17
24	69 14	37 14	57 51	29 83	04 74	39 97	01 07	06 00	39 42	25 67
25	93 26	85 88	07 39	99 74	79 93	23 57	04 91	26 67	28 85	32 69

Twenty-eighth Thousand

	1-4	5-8	9-12	13-26	17-20	21-24	25-28	29-32	33-36	37-40
1	28 71	02 23	<u>43 06</u>	91 65	64 38	67 77	46 71	23 60	33 82	26 86
2	87 67	68 27	92 76	84 27	78 05	11 45	03 53	58 61	18 34	62 83
3	46 91	72 15	<u>01 66</u>	77 04	91 15	15 90	71 30	71 13	36 46	55 79
4	51 58	01 47	38 99	04 13	06 63	38 96	21 00	35 16	71 69	09 34
5	82 57	97 55	35 08	40 72	14 70	14 32	03 62	26 88	50 36	63 17
6	39 81	59 34	26 82	53 26	93 19	31 50	84 19	84 25	11 93	51 32
7	31 63	54 73	89 35	21 86	68 66	42 16	38 01	92 58	57 90	04 77
8	74 07	21 92	71 39	64 30	64 59	96 61	49 31	20 13	96 55	42 26
9	69 88	92 27	07 97	56 18	07 69	33 34	54 87	10 07	77 20	23 60
10	88 58	54 78	53 87	18 10	01 92	28 97	87 42	91 22	53 34	69 24
11	41 55	82 54	18 66	29 63	16 57	78 73	43 21	34 37	29 30	31 19
12	14 04	73 77	79 63	81 90	53 48	45 02	00 11	88 84	97 24	88 92
13	57 23	32 53	61 75	83 74	35 97	93 40	18 28	18 28	14 27	73 57
14	77 86	75 66	41 14	86 92	09 49	84 49	56 55	65 74	56 35	73 45
15	66 31	76 91	37 52	49 53	60 23	12 62	12 95	41 09	88 61	34 45
16	79 94	75 87	09 06	50 70	70 54	00 83	74 50	86 81	32 54	19 62
17	80 81	33 38	25 79	84 58	51 65	04 41	45 12	29 84	09 04	25 89
18	92 80	74 72	47 57	18 91	72 35	18 77	41 23	86 21	54 25	52 84
19	90 19	62 76	05 92	76 00	51 00	54 67	23 52	81 76	73 49	67 72
20	53 46	38 68	95 41	92 41	00 87	78 40	31 94	42 78	42 85	17 96
21	03 68	12 51	57 25	29 30	08 55	97 02	36 66	44 25	58 56	60 16
22	45 50	95 74	02 36	33 14	75 50	52 12	62 58	33 72	09 63	24 19
23	39 56	53 09	35 17	97 19	92 49	62 27	67 91	84 52	60 12	66 40
24	18 81	93 07	70 29	59 67	72 26	92 11	57 29	39 89	82 12	00 56
25	22 42	08 10	73 13	60 26	77 32	82 09	42 09	30 64	50 90	55 37

* Reproduced from page 17 of *TRACTS FOR COMPUTERS XXIV*, Cambridge University Press, 1939, by permission of Professor E.S. Pearson.

REFERENCES

1. COCHRAN, WILLIAM G,
Sampling Techniques (Tokyo : John Willey & Son, 1972).
2. FREUD, JOHN E and FRANK J. WILLIAM,
Modern Business Statistics (Tokyo : Charles E. Tuttle co, 1965),
Chaps 6, 7, 8, 12.
3. GOODE, WILLIAM J and PAUL K. HATT,
Methods in Social Research (Tokyo : Mac Graw Hill Co., 1952), Chaps 14.
4. LOTT, RICHARD W.,
Basic Data Processing (New Jersey : Prentice Hall, 1967), Chaps 1.
5. MILLS, FREDERICK C.,
Introduction to Statistics (New York : Henry Holt Co., 1956).
Chap 2; Chap 3;
6. NOEGROHO,
Sendi-sendik Statistik (Jakarta : P.T. Pembangunan, 1963), Bab VIII.
7. WILSON, E. BRIGHT,
An Introduction to Scientific Research, (New York : Mac Graw Hill Co,
1952), Chap 7.

6. PENGOLAHAN DATA

Data yang telah terkumpul belum bisa dibaca, sebab belum dapat secara langsung menggambarkan karakteristik populasi yang diteliti. Supaya dapat dengan mudah membacanya data yang telah dikumpulkan itu harus diolah lebih lanjut. Inilah yang disebut pengolahan data (*data processing*).

Pengolahan data ini memiliki dua arti. Dalam arti luas, pengolahan data meliputi setiap langkah yang diambil, bagaimana pun caranya, agar data tersebut dapat digunakan untuk maksud tertentu.

Dalam arti yang lebih sempit, pengolahan data meliputi pengetikan dan pengolahan oleh komputer elektronik.

Yang dimaksud dengan pengolahan data dalam buku ini ialah pengolahan data dalam arti luas, yaitu mengolah data yang cerai-berai, dikumpulkan secara teratur, diurut, sehingga dapat dengan lebih mudah diketahui karakteristik populasi yang kita selidiki.

Misalnya nilai ujian biologi satu kelompok mahasiswa, yang diadakan di IKIP Surabaya 1975 adalah sebagai berikut:

32	51	61	32	49	65	35	76	53	58
69	31	56	77	63	43	81	33	58	28
44	57	25	61	50	96	60	36	78	74
60	89	53	59	55	54	60	80	38	70
75	37	47	64	40	71	73	63	72	54
29	51	56	62	57	61	39	53	36	35
62	34	45	45	33	46	13	41	50	65
38	68	37	48	59	58	58	49	57	41
95	09	66	60	27	39	46	46	14	55
56	83	34	62	72	55	44	64	42	79

Data (nilai) ini masih belum disusun teratur. Masih belum jelas benar berapa orang yang mendapat nilai 80 atau lebih. Dan sukar dibaca berapa prang yang mendapat nilai di bawah 30. Oleh karena itu, untuk memudahkan pembacaan, data itu perlu disusun lebih dahulu ke dalam satu tabel. Ada beberapa bentuk tabel yang dapat dibuat peneliti bila ia ingin memasukkan nilai di atas ke dalam sebuah tabel, antara lain seperti contoh berikut:

(1) Contoh 1.

TABEL NILAI UJIAN BIOLOGI MAHASISWA IKIP SURABAYA 1975

NILAI	JUMLAH MAHASISWA
5 - 9	1
10 - 14	2
15 - 19	0
20 - 24	0
25 - 29	4
30 - 34	7
35 - 39	10
40 - 44	7
45 - 49	9
50 - 54	9
55 - 59	15
60 - 64	14
65 - 69	5
70 - 74	6
75 - 79	5
80 - 84	3
85 - 89	1
90 - 94	0
95 - 99	2

(2) Contoh 2.

TABEL NILAI UJIAN BIOLOGI MAHASISWA IKIP SURABAYA 1975

NILAI	JUMLAH MAHASISWA
5 - 39	24
40 - 74	65
75 -109	11

Cobalah bandingkan kedua contoh tabel di atas! Kita dapat melihat bahwa :

- (a) tabel contoh 1 sangat panjang, sedang tabel contoh 2 sangat pendek.
- (b) kedua tabel itu benar, tetapi tidak umum dipakai orang.
- (c) kedua tabel itu lebih mudah dibaca, dibandingkan dengan nilai-nilai yang masih belum disusun secara teratur.

Untuk menghindarkan agar tabel tidak terlalu panjang dan tidak terlalu pendek, umumnya digunakan rumus Sturges sebagai pedoman. Rumus ini digunakan untuk menentukan banyaknya kelas (k) yang dapat dihitung sebagai berikut :

$$k = 1 + 3,3 \log n.$$

$$k = \text{banyaknya kelas}$$

$$n = \text{banyaknya sampel data}$$

Kalau nilai-nilai di atas yang banyaknya 100 kita tentukan banyaknya kelas dalam satu tabel, maka hasilnya adalah :

$$\begin{aligned} k &= 1 + 3,3 \log 100 \\ &= 1 + 3,3 \times 2 = 7,6 \end{aligned}$$

Kelas-kelas dalam tabel jumlahnya selalu bilangan bulat. Karena itu kita boleh menggunakan 7 kelas atau 8 kelas. Sehingga tabel yang baik adalah sebagai berikut :

TABEL NILAI UJIAN BIOLOGI MAHASISWA IKIP SURABAYA 1975

NILAI	f
0 - 14,9	3
15 - 29,9	4
30 - 44,9	24
45 - 59,9	33
60 - 74,9	25
75 - 89,9	9
90 - 104,9	2

Dalam praktek, f (frekuensi) diperoleh dengan cara membuat goresan pada waktu kita menghitung nilai-nilai yang terletak dalam masing-masing kelas, misalnya :

NILAI	GORESAN
0 - 14,9	///
15 - 29,9	////
30 - 44,9	#### #### #### #### ////
45 - 59,9	#### #### #### #### #### #### ///
60 - 74,9	#### #### #### #### ####
75 - 89,9	#### ////
90 - 104,9	//

Walaupun banyaknya kelas menurut rumus Sturges dapat dihitung, namun banyaknya kelas belum merupakan jaminan sebagai tabel yang baik. Rumus Sturges belum menjawab :

(a) Apakah selalu batas bawah kelas terkecil itu 0 ?

Jawabnya, tidak. Boleh dikira-kira sendiri.

(b) Bila kelasnya tidak 7, tetapi 8, apakah rata-ratanya akan sama?

Jawabnya, tidak selalu sama. Bisa berbeda sedikit. Sebab nilai statistik adalah nilai perkiraan. Berbeda dengan matematik.

Rata-rata bila dihitung dari nilai yang cerai-berai juga berbeda dengan bila nilai tersebut telah dikelompokkan ke dalam kelas-kelas (tabel).

Latihan-latihan dan pengalaman membuat tabel sangat membantu pekerjaan membuat tabel yang baik. Pedoman untuk membuat tabel yang baik adalah sebagai berikut :

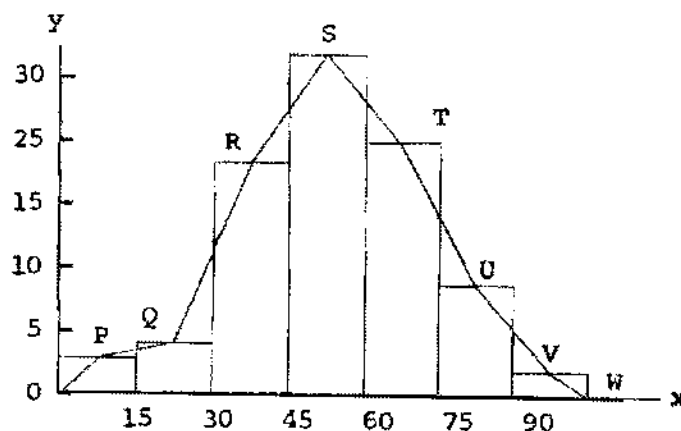
- (1) besar kelas adalah sama
- (2) batas kelas tidak boleh sama
- (3) tidak boleh menggunakan kelas terbuka.

Contoh tabel dengan susunan kelas yang tidak baik ialah :

TABEL NILAI UJIAN STATISTIK MAHASISWA IKIP SURABAYA 1975

NILAI	f
2 - 8	6
8 - 24	20
24 - 40	32
40 - 65	25
65 - ke atas	47

Cobalah perhatikan di mana kesalahan tabel di atas ini? Bila ke 100 nilai itu disajikan dalam bentuk histogram, kita akan memperoleh gambar sebagai berikut :

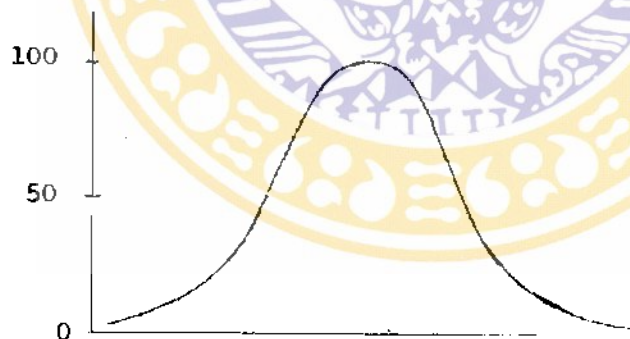


Histogram nilai ujian Biologi

Garis yang menghubungkan titik-titik tengah sisi atas sebuah histogram (perhatikan garis OPQRSTUVW) membentuk suatu kurva yang disebut kurva bentuk poligon (*polygon*).

Bagaimana bentuk kurvanya bila tidak hanya 100 mahasiswa angkatan 1975 IKIP Surabaya saja yang nilai ujian biologinya diplot dalam bentuk kurva? Jika nilai-nilai ujian biologi untuk semua angkatan yang sudah menempuh ujian dan yang akan menempuh ujian diplot data-nya, tentu kita akan memperoleh data (nilai) yang *infinite*. Akan menjadi bagaimana rupa dan bentuk kurva?

Secara statistik kita menarik kesimpulan bahwa pencaran nilai ujian biologi (semua angkatan mahasiswa IKIP Surabaya yang pernah dan yang akan menempuh ujian adalah seperti pencaran nilai sampelnya (100 mahasiswa)). Hanya bedanya bahwa pencaran nilai untuk populasi yang *infinite* kurvanya berupa garis khayal berbentuk genta. Kurvanya dapat digambarkan sebagai berikut :



Kurva nilai ujian biologi mahasiswa
IKIP Surabaya

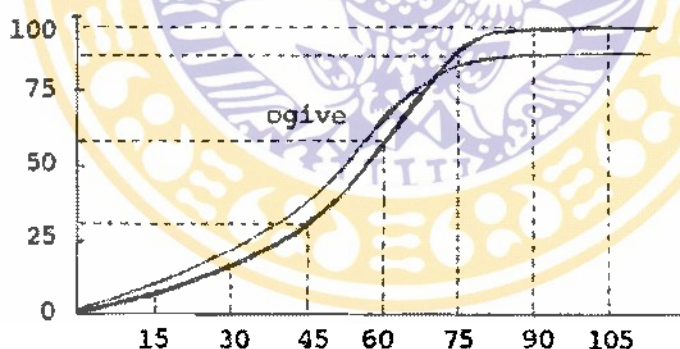
Kurva (*curve*) ini adalah bayangan dalam otak kita tentang distribusi frekuensi nilai biologi mahasiswa IKIP (populasi), yang bertolak dari hasil pengamatan dan pengukuran nilai sampel.

Data tentang nilai ujian biologi seperti tabel di atas dapat juga disajikan sebagai berikut :

TABEL NILAI UJIAN BIOLOGI MAHASISWA IKIP SURABAYA 1975

NILAI		f. kum
Kurang dari	0	0
Kurang dari	15	3
Kurang dari	30	7
Kurang dari	45	31
Kurang dari	60	64
Kurang dari	75	89
Kurang dari	90	98
Kurang dari	105	100

Penyajian data seperti ini menggambarkan distribusi frekuensi disebut distribusi frekuensi kumulatif (f.kum.). Kurva distribusi frekuensi kumulatif ini disebut *ogive*.



Garis khayal, yang licin, yang menjadi resultante dari ogive itu disebut kurva juga.

Data penelitian tidak selalu bersifat kuantitatif. Lebih-lebih kalau datanya berasal dari penelitian Ilmu Sosial. Kadang-kadang timbul problema untuk mentransfer data kualitatif yang disebut atribut ke data kuantitatif yang serial, yang disebut satu variabel. Problemanya ialah bagaimana caranya menyatakan data kualitatif ke kuantitatif untuk variabel :

(a) partisipasi sosial

- (b) latar belakang keluarga
- (c) pendidikannya
- (d) status sosial dan sebagainya.

Data penelitian sosial seperti contoh di atas biasanya disajikan dalam bentuk skala.

Teknik skala (*scaling techniques*) ada beberapa macam, diantaranya ialah skala untuk mengukur jarak sosial (*social distance*) dan *rating scale*.

(1) Skala jarak sosial.

Jarak sosial (*social distance*) itu ialah kadar atau derajad pengertian dan keberatan yang memberi corak hubungan sosial pada umumnya.

Bogardus mengukur jarak sosial 1725 orang Amerika terhadap empat golongan minoritas di Amerika. Teknik mengukur dan hasil pengukurannya disajikan dalam tabel di bawah ini.

	(1) To close kindship by marriage	(2) To my club as person- al chums	(3) To my street as neigh- bors	(4) To em- ploy- ment in my occu- pation	(5) To citi- zenship in my country	(6) As visi- tors only to my country	(7) Would exclude from my country
Regarding eth- nic mi- norities listed							
English	93.7	96.7	97.3	95.4	95.9	1.7	0.0
Swedish	45.3	62.1	75.6	78.0	86.3	5.4	1.0
Polish	11.0	11.6	28.3	44.3	58.3	19.7	4.7
Korean	1.1	6.8	13.0	21.4	23.7	47.1	19.1

Adapted from Emory S. Bogardus, *Immigration and Race Attitudes*
(Boston : Heath, 1928), p. 25.

Pada tabel itu disajikan beberapa seri kelompok menurut *ethnis* dan yang masing-masing seri dibagi dalam 7 skala. Berdasarkan asumsi ke 7 atribut tersebut disusun dalam skala kontinum dari jarak-jarak sosial. Dari tabel tersebut kita melihat bahwa skala 6 dan 7 tidak termasuk dalam satu kontinum jika dibanding dengan skala 1 - 5. Kenyataan ini terlihat jika data pada tabel tersebut dinyatakan dalam bentuk kurva seperti terlihat pada gambar berikut:

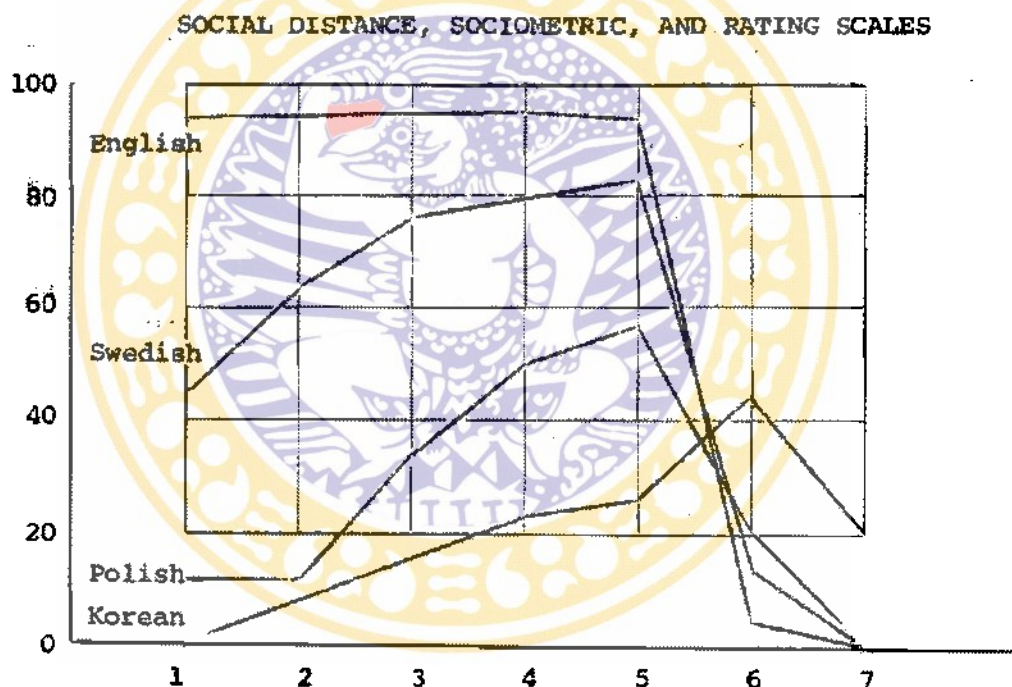


FIGURE 1. Frequency of responses of 1,725 Americans to four different ethnic minorities.

Setelah skala 5 garis kurjanya jatuh dengan sudut yang tajam sekali. Jika respons 6 dan 7 (skala 6 dan 7) dihilangkan kita memperoleh kurva seperti terlihat pada gambar berikut :

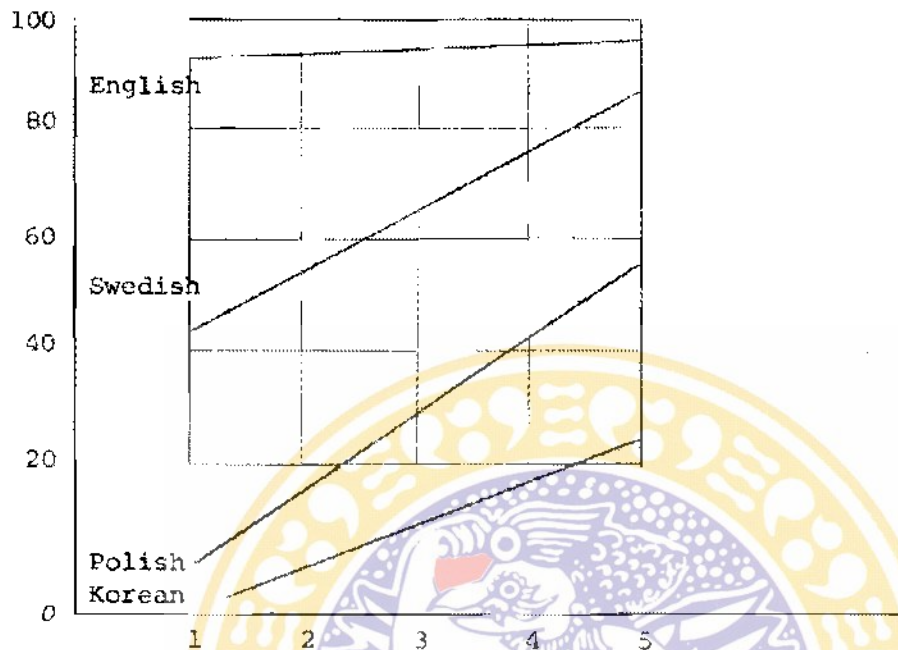


FIGURE 2. Smoothed curves of social-distance score for four ethnic minorities.

Indeks yang dipakai untuk menyatakan posisi setiap kelompok minoritas itu ialah :

- (i) tingginya garis,
- (ii) miringnya (slope) garis, yaitu sudut garis miring itu dengan garis vertikal.

Perhatikan garis (jarak-sosial) minoritas Inggris dan Polandia terhadap masyarakat Amerika. Inggris menduduki (mencerminkan) posisi yang lebih menguntungkan dari pada Polandia sebab :

1. posisi garis Inggris letaknya lebih tinggi
2. jarak kedua kurva (Inggris dan Polandia) jauh lebih lebar pada posisi 1 dari pada posisi 5.

Metode yang paling praktis untuk mengukur jarak-sosial seperti ini ialah dengan jalan mengalikan prosentasi (%) masing-masing kelompok minoritas dengan posisinya. Hasil perkalian itu menjadi angka (score) kelompok tersebut.

Contohnya dapat dilihat pada tabel di bawah ini.

TABLE CALCULATION OF WEIGHTED SOCIAL-DISTANCE SCORES FOR FOUR ETHNIC MINORITIES

POSITION (1)	GROUPS							
	English		Swedish		Polish		Korean	
	(2) Per Cent	(3) [(1) x (2)]	(4) Per Cent	(5) [(1) x (4)]	(6) Per Cent	(7) [(1) x (6)]	(8) Per Cent	(9) [(1) x (8)]
1	93.7	93.7	45.3	45.3	11.0	11.0	1.1	1.1
2	96.7	193.4	62.1	124.2	11.6	23.2	10.8	21.6
3	97.3	291.9	75.6	226.8	28.3	84.9	11.8	35.4
4	95.4	381.6	78.0	312.0	44.3	177.2	20.1	80.4
5	95.9	479.5	86.3	431.5	58.3	291.5	27.5	137.5
Total		1,440.1		1,139.8		587.8		276.0

Dari tabel tersebut kita dapat melihat bahwa jarak sosial golongan minoritas tersebut adalah dari 1440 sampai 276.

Kelemahan perhitungan jarak-jarak sosial dari Bagordus ini ialah :

- a) karena Bagordus berasumsi bahwa skala jarak-sosial selalu sama untuk titik-titik yang berurutan dalam skala. Kenyataannya tidak demikian
- b) tidak ada titik nol dalam skala jarak sosial. Akibatnya, kita tidak mungkin mengetahui jarak sosial Swedia atau Polandia. Apakah jarak sosial Polandia dua kali dari Korea dan Swedia dua kali dari Polandia?

(2) Pengukuran Sosiometrik.

J.L. Mareno dan Helen Hall Jennings mengukur jarak sosial dengan cara yang sangat berbeda dengan cara Bagordus. Cara Helen-Jennings, yang disebut sosiometri, adalah cara mengukur penarikan penolakan antar individu dalam satu kelompok kecil. Teknik sosiometri ini digunakan pada berbagai situasi misalnya

Untuk kelompok informal, kelas-kelas di sekolah, penjara dan industri. Contoh matriks sosiometri yang menyatakan "siapa memilih siapa" antar anggota panitia sekolah dapat dilihat pada gambar berikut.

		Chosen															
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Chooser	A		3	2				1									
	B	2				1	3										
	C																
	D		3	1			2										
	E			3				2									1
	F				2			1				3					
	G	3								2		1					
	H																
	I					3		1				2					
	J				3		2	1									
	K		1					2		3							
	L	3								1						2	
	M								2	1						3	
	N							1		3							2
	O				2			1				3					
	P							2				3				1	
1 st Choice			1	1		1		6		2		1				1	1
2 nd Choice		1		1	2		2	3	1	1		1				1	1
3 rd Choice		2	2	1	1	1	1			2		3				1	
Total		3	3	3	3	2	3	9	1	5		5				3	2

Data sosiometri dari matriks tersebut dapat disajikan dalam satu *sosiogram* seperti gambar berikut.

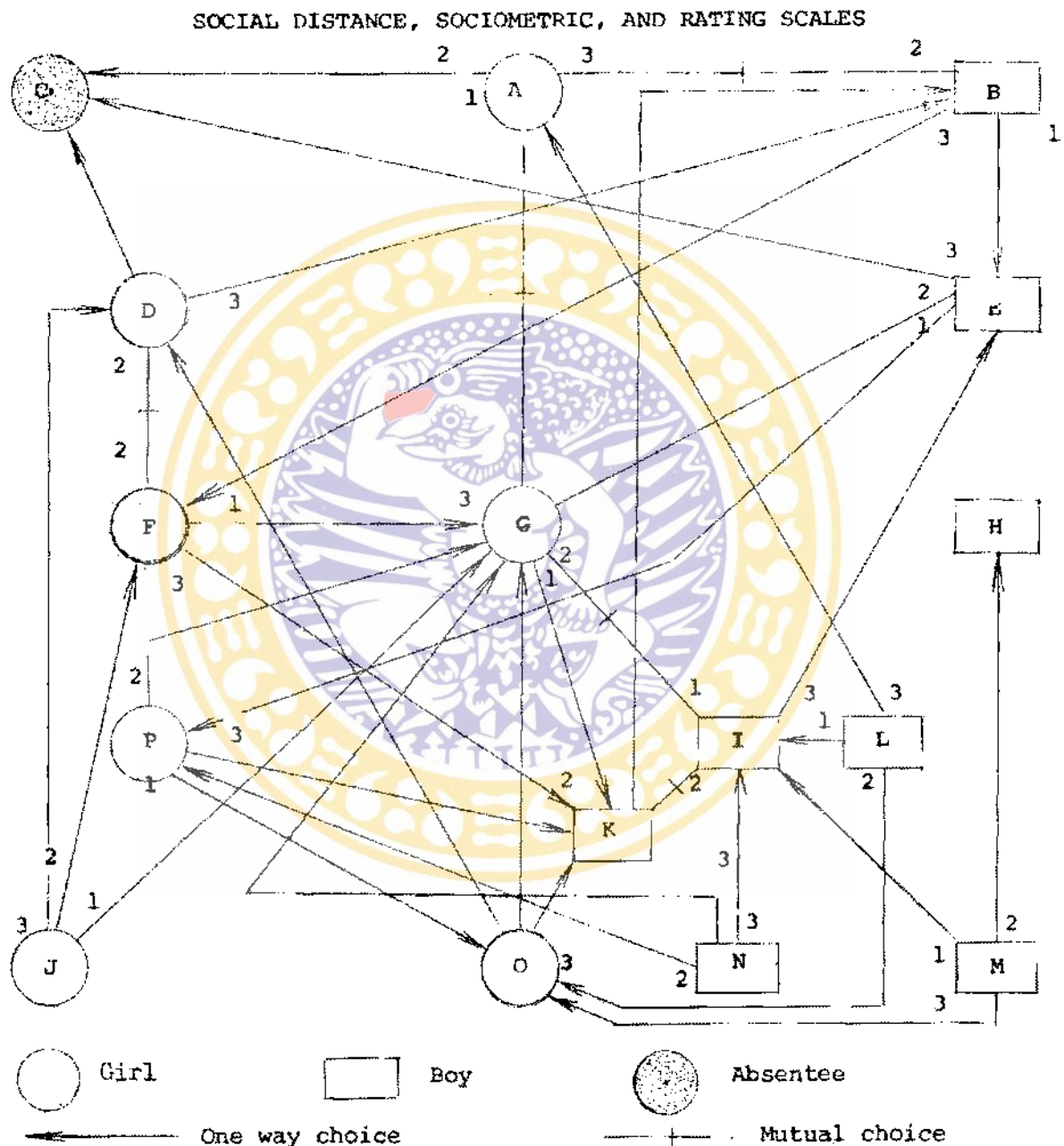


FIGURE 4. Example of a sociogram, presenting data from matrix in Figure 3.

Cobalah berusaha membaca dan memahami data penelitian dalam berbagai bentuk penyajiannya seperti di atas. Penyajian data dalam bentuk tabel, grafik, matriks, dan sosiogram sangat menyherhanakan membaca masalah yang kompleks.

LAMPIRAN IV

MODEL USUL PROYEK PENELITIAN

1. Judul Usul Proyek Penelitian :
2. Pemimpin Proyek Penelitian :
- a. Nama lengkap :
- b. Pangkat dan Jabatan :
- c. Tempat Penelitian (Alamat) :
- c.1. Laboratorium/Seksi :
- c.2. Bagian/Jurusan :
- c.3. Fakultas :
3. Jangka Waktu Penelitian : bulan.
- Mulai :
Selesai :
4. Biaya Yang Diperlukan : Rp.
(.....)
(.....)

Menyetujui:

Pemimpin Proyek Penelitian

Dekan Fakultas

Universitas/Institut :

(.....)

(.....)

Mengetahui:

Pimpinan Universitas/Institut

Nama Lengkap :

Pangkat dan Jabatan:

Tanda Tangan :

LAPORAN PENELITIAN

PENGETAHUAN DASAR PENELITIAN SAN STATISTIK

JAWA TIMUR

A. Judul Penelitian :

B. Ruang lingkup/Bidang Pengetahuan :

C. Intisari/Abstract :

(Uraikan dengan singkat tujuan penelitian dan cara pelaksanaannya).

D. Latar Belakang :

(Jelaskan tentang pentingnya penelitian yang diusulkan dan memang belum terjawab. Di samping itu dapat pula diuraikan tentang guna hasil (Significance) yang diharapkan).

E. Penelaahan Study dan Kepustakaan :

(Uraikan penelitian-penelitian yang telah pernah ada dan yang mempunyai kaitan dengan usul proyek penelitian serta tunjukkan bahwa masalah tersebut belum secara tuntas pernah diselidiki).

F. Tujuan Penelitian :

G. Metodologi :

(Uraikan dengan jelas dan lengkap metodologi yang digunakan)

H. Jadwal Pelaksanaan :

(Kalau dapat berupa satu sketsa tentang Jadwal Waktu dari Management Penelitian).

I. Personalia dan Organisasi :

I.1. Pemimpin Proyek :

1.1. Nama lengkap :

- 1.2. Pangkat dan Jabatan :
- 1.3. Tempat penelitian/alamat:
- 1.4. Tugas/Hubungan kerja :
- 1.5. Waktu yang disediakan :
untuk penelitian ini
(dalam jam per minggu)

- I.2. **Pelaksana** Proyek :
- 2.1. Nama lengkap :
- 2.2. Pangkat dan Jabatan :
- 2.3. Alamat kerja :
- 2.4. Tugas/Hubungan kerja :
- 2.5. Waktu yang disediakan
untuk penelitian ini :

Kalau lebih dari 1 (satu) orang pelaksana proyek nyatakan dengan Pelaksana I, Pelaksana II, dst.

- I.3. Teknisi/Tenaga Ahli/Pembantu:
Pelaksana
(sama seperti di atas)
- I.4. Tenaga Administrasi :
(berapa orang)
- I.5. Pekerja Lapangan :
(berapa orang kalau dapat
diperinci tugasnya masing-
masing.

Catatan:

Pengalaman, baik di bidang penelitian maupun di bidang pendidikan dari Pimpinan Proyek, Pelaksana Proyek dan Pembantu Pelaksana diuraikan secara ringkas tetapi lengkap dalam lampiran Curricula Vitae.

J. Biaya Penelitian :

(dalam waktu maksimum 1 (satu) tahun).

Rekapitulasi:

J.1. Upah:

Personalia : Jam/minggu : Jumlah honorarium
 Jumlah (Rp.):
 (selama proyek) (per jenis biaya)

- 1.1. Pimpinan proyek
- 1.2. Pelaksana I
- 1.3. Pelaksana II, dst.
- 1.4. Pembantu Pelaksana I
- 1.5. Pembantu Pelaksana II, dst.
- 1.6. Tenaga Administrasi
- 1.7. Pekerja Lapangan.

J.2. Alat-alat :

Spesifikasi : Jumlah satuan : Harga/satuan : Jumlah (Rp.)
 : Jumlah (Rp.) : (per jenis biaya)

- 2.1.
- 2.2
- 2.3. dst.

J.3. Bahan-bahan :

Spesifikasi : Jumlah satuan : Harga/satuan : Jumlah (Rp.)
 : Jumlah (Rp.) : (per jenis biaya)

- 3.1.
- 3.2.
- 3.3. dst.

J.4. Perjalanan :

- 4.1. Perjalanan lokal : Jumlah (Rp.) : Jumlah (Rp.)
 (per jenis biaya)
 - 4.2. Perjalanan dari/ke :
-

I.12e. Tenaga-tenaga Pembantu.

Tenaga-tenaga pembantu adalah sekalian tenaga yang membantu para tenaga peneliti dalam penyelenggaraan kegiatan-kegiatan penelitian yang bersangkutan, seperti pengedar daftar isian pertanyaan, bendaharawan, dan sebagainya.

